KU LEUVEN

ARENBERG DOCTORAL SCHOOL
Faculty of Science

UNIVERSITEIT
iYUNIVESITHI
STELLENBOSCH
UNIVERSITY

STELLENBOSCH UNIVERSITY
Faculty of Economic and Management Sciences

Extreme Quantile Inference

Sven Buitendag

Supervisors:
Prof. dr. Jan Beirlant
Prof. dr. Tertius de Wet

Dissertation presented in partial fulfillment of the requirements for the joint degree of Doctor of Science (PhD) in Mathematics at KU Leuven and Doctor of Philosophy (PhD) in Mathematical Statistics at Stellenbosch University

March 2020

Stellenbosch University https://scholar.sun.ac.za

Extreme Quantile Inference

Sven BUITENDAG

Examination committee:
Prof. dr. Wim Schoutens, chair
Prof. dr. Jan Beirlant, supervisor
Prof. dr. Tertius de Wet, supervisor
Prof. dr. An Carbonez
Prof. dr. Tim Verdonck
Prof. dr. Ivette Gomes
(University of Lisbon)
Prof. dr. Hansjörg Albrecher
(HEC Lausanne)
Dr. Francois Kamper
(Stellenbosch University)

Dissertation presented in partial fulfillment of the requirements for the joint degree of Doctor of Science (PhD) in Mathematics at KU Leuven and Doctor of Philosophy (PhD) in Mathematical Statistics at Stellenbosch University

March 2020

Stellenbosch University https://scholar.sun.ac.za

Copyright © 2020 Stellenbosch University
All rights reserved

© 2019 KU Leuven – Faculty of Science
Uitgegeven in eigen beheer, Sven Buitendag, Celestijnenlaan 200A box 2402, B-3001 Leuven (Belgium)

Alle rechten voorbehouden. Niets uit deze uitgave mag worden vermenigvuldigd en/of openbaar gemaakt worden door middel van druk, fotokopie, microfilm, elektronisch of op welke andere wijze ook zonder voorafgaande schriftelijke toestemming van de uitgever.

All rights reserved. No part of the publication may be reproduced in any form by print, photoprint, microfilm, electronic or any other means without written permission from the publisher.

Stellenbosch University https://scholar.sun.ac.za

Acknowledgements

I would firstly like to thank my two supervisors, Prof Tertius de Wet and Prof Jan Beirlant, for their invaluable support and contributions during my doctoral studies on both an academic and personal level. Their deep insights into statistics, the sheer pleasure of working with them and the openness in which we engage with one another has all contributed to the research that we have generated and personal relationships that we have established. I deeply appreciate all of your efforts during my doctoral studies, and for making research and statistics a pleasure.

I am grateful to Graziella del Savio and Dorothy Stevens for their help in drafting and finalizing my joint doctoral agreement, to KU Leuven, Stellenbosch University and the NRF for funding my doctoral studies and travels, and to the jury members Francois Kamper, Hansjöerg Albrecher, Ivette Gomes, Wim Schoutens, An Carbonez and Tim Verdonck for their time and effort in assessing this dissertation.

I would also like to thank my mother, brother and sister for their help and support during my doctoral studies. Their calls and messages meant a great to me during my stay in Belgium, and I remember their visits with sincere fondness.

This work is dedicated to my wife Elena, whose love, wisdom and energy is the substance with which this research came to fruition. Thank you for your wonder each and every day.

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Abstract

A novel approach to performing extreme quantile inference is proposed by applying ridge regression and the saddlepoint approximation to results in extreme value theory. To this end, ridge regression is applied to the log differences of the largest sample quantiles to obtain a bias-reduced estimator of the extreme value index, which is a parameter in extreme value theory that plays a central role in the estimation of extreme quantiles. The utility of the ridge regression estimators for the extreme value index is illustrated by means of simulations results and applications to daily wind speeds.

A new pivotal quantity is then proposed with which a set of novel asymptotic confidence intervals for extreme quantiles are obtained. The ridge regression estimator for the extreme value index is combined with the proposed pivotal quantity together with the saddlepoint approximation to yield a set of confidence intervals that are accurate and narrow. The utility of these confidence intervals are illustrated by means of simulation results and applications to Belgian reinsurance data.

Multivariate generalizations of sample quantiles are considered with the aim of developing multivariate risk measures, including maximum correlation risk measures and an estimator for the extreme value index. These multivariate sample quantiles are called center-outward quantiles, and are defined as an optimal transportation of the uniformly distributed points in the unit ball $\mathbb{S}^d$ to the observed sample points in $\mathbb{R}^d$. A continuous extension of the center-outward quantile is proposed, which yields quantile contours that are nested. Furthermore, maximum correlation risk measures for multivariate samples are presented, as well as an estimator for the extreme value index for multivariate regularly varying samples. These results are applied to Danish fire insurance data and the stock returns of Google and Apple share prices to illustrate their utility.

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Contents

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Chapter 1

Introduction

This introductory chapter serves to provide context in which extreme quantile inference plays a crucial role in understanding the behavior of extreme events. We hence provide an overview of the mathematical theory underlying extreme quantile inference and briefly outline the topics on which this thesis focuses. To this end this chapter is structured around three topics: the background as to the role of extreme value analysis in the statistical understanding of extreme events, an overview of extreme value theory, next to the underlying methodology of extreme quantile inference, and finally a brief outline of this thesis.

1.1 Background

The statistical analysis of extreme events, referred to as *extreme value analysis*, plays a crucial role in fields of study charged with understanding and managing such events, including meteorology, seismology, hydrology, engineering, finance, insurance among others. Extreme events are defined as rare events which have large impact, such as stock returns during financial crises, wind speeds during storms, claim amounts paid by reinsurers, the carat size of large diamonds, or the damage caused by natural disasters. Thanks to the mathematical results of Fréchet (1927), Fisher and Tippett (1928), von Mises (1936), and Gnedenko (1943) the statistical behavior of extreme events are mathematically well understood, and as a result the statistical methodology developed to quantify the behavior of extreme events.

Stellenbosch University https://scholar.sun.ac.za

Central to this theory is a shape parameter called the *extreme value index* (EVI) which is a measure of the tail-heaviness of the underlying distribution. The EVI features in the estimation of many important extremal statistical quantities, including extreme quantiles and returns periods. Estimation of the EVI, extreme quantiles and return periods has received much attention, including Embrechts et al. (1997), Beirlant et al. (2004), and de Haan and Ferreira (2006), to name a few general textbooks.

The applications of extreme value theory translates into practical risk management tools, such as the calculation of the height of a dyke that should withstand a once-in-10000-years storm, or the amount of cash reserves a bank should hold to withstand a once-in-100-years financial shock, or the decision which cities should be evacuated in preparation of a once-in-500-years storm, or the calculation of the price a reinsurance company should ask to expect to remain profitable.

1.2 Overview

Let $X_1, X_2, \ldots, X_n$ denote a sample of n independent identically distributed random variables with cumulative distribution function F and quantile function Q defined as the left continuous inverse function of F. The ordered data are denoted as $X_{1,n} \leq X_{2,n} \leq \cdots \leq X_{n,n}$. Our principle focus is on extreme quantiles $Q(1-p)$ for p close to 0, which one would intuitively infer about using the extreme sample quantiles $X_{n,n}, X_{n-1,n}, \ldots, X_{n-k,n}$ for k close to n. To this end we first consider the asymptotic theory of the sample maximum $X_{n,n}$, which in turn leads to an appropriate setting to study tail properties.

The asymptotic distribution theory of the sample maximum $X_{n,n}$ concerns two questions: do there exist real sequences $a_n > 0$ and b_n such that $P(a_n^{-1}(X_{n,n} - b_n) \leq x) \to G(x)$ as $n \to \infty$ for some non-degenerate distribution function G (called the extremal limit problem), and, determine the class of distributions F for which this limit in distribution to G holds (called the domain of attraction conditions).

Fisher and Tippett (1928) and Gnedenko (1943) proved that the distribution function G necessarily belongs to the class of generalized extreme value (GEV)

Stellenbosch University https://scholar.sun.ac.za

distributions $\{G_\gamma, \gamma \in \mathbb{R}\}$ which have the following form:

$$G_\gamma(x) = \begin{cases} \exp\left(-(1+\gamma\, x)^{-1/\gamma}\right) & \text{for } x > -1/\gamma \quad \text{when } \gamma \neq 0, \\ \exp\left(-\exp\left(-x\right)\right) & \text{for } x \in \mathbb{R} \quad \text{when } \gamma = 0. \end{cases} \tag{1.1}$$

The shape parameter $\gamma \in \mathbb{R}$ is called the EVI.

The domain of attraction condition on F is then specified as follows:
there exist real sequences $a_n > 0$ and b_n such that $P(a_n^{-1}\ (X_{n,n} - b_n) \leq x) \to G_\gamma(x)$ as $n \to \infty$ if and only if

$$\frac{T(xu) - T(x)}{a(x)} \to h_\gamma(u) \quad \text{for} \quad u > 0 \quad \text{as} \quad x \to \infty \tag{1.2}$$

where $a(x)$ is a positive function, $h_\gamma(u) = \begin{cases} \frac{u^\gamma - 1}{\gamma} & \text{if } \gamma \neq 0 \\ \log u & \text{if } \gamma = 0 \end{cases}$, and $T(x) = Q\left(1 - \frac{1}{x}\right)$ is called the tail quantile function.

1.2.1 Pareto-type Distributions

The distributions which satisfy the domain of attraction condition for a positive EVI $\gamma > 0$ are called Pareto-type distribution functions. The distributions from this class have heavy tails, such as the Student-t, Pareto and Fréchet distributions. For $\gamma > 1$ the mean of the underlying distribution is infinite, and for $\gamma > 1/2$ the variance of the underlying distribution is infinite. Financial and non-life insurance data often exhibit Pareto-type behavior, as discussed for instance in Embrechts et al. (1997) and Beirlant et al. (2004).

Pareto-type distribution functions have a survival function that can be written as $1 - F(x) = x^{-1/\gamma}\, \ell_F(x)$, or equivalently, have a tail quantile function that can be written as $T(x) = x^\gamma\, \ell_T(x)$ where the functions ℓ_T and ℓ_F are slowly-varying:

$$\ell(ty)/\ell(t) \to 1 \text{ as } t \to \infty \text{ for all } y > 1.$$

It was derived under general conditions on ℓ_F or ℓ_T that the weighted log-spacings $Z_j = j\ (\log X_{n-j+1,n} - \log X_{n-j,n})$ have an asymptotic exponential distribution with mean γ:

$$Z_j = \gamma\, E_j\ + o_p(1) \quad \text{for } j = 1, 2, \ldots, k \tag{1.3}$$

Stellenbosch University https://scholar.sun.ac.za

as $k, n \to \infty$ such that $k/n \to 0$ where $E_1, E_2, \ldots, E_n$ are independent standard exponential random variables (see for instance Chapter 3 in Beirlant et al. (2004)).

If the slowly varying function ℓ_F is constant then the underlying distribution is a Pareto distribution with tail index $\alpha = 1/\gamma$ and the weighted log-spacings would be exactly exponentially distributed with mean γ. The deviation of the slowly varying part ℓ_F from a constant therefore translates into how far the underlying distribution differs from a strict Pareto distribution. This also provides a way to investigate whether a random sample originates from a Pareto-type distribution: if $\gamma > 0$ then the Pareto QQ-plot

$$\left(-\log(1 - \frac{i}{n+1}), \log X_{i,n} \right), \ i = 1, \ldots, n,$$

has a gradient approximately equal to the EVI γ at the largest sample quantiles.

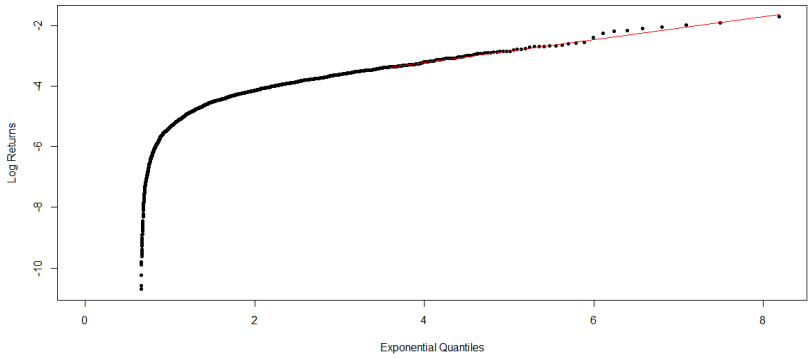

Figure 1.1: A Pareto QQ-plot of the daily log-returns of Google stock from 2005 to 2019 with a sample size of 3600. The red line illustrates the gradient of the hundred largest points, and can be used to estimate the EVI.

One of the most popular estimators for $\gamma > 0$ is the Hill (1975) estimator which is motivated by the tail of the Pareto QQ-plot as illustrated in Figure 1.1, as well as by the weighted log-spacings' representation from (1.3). This estimator is defined as

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} = \frac{1}{k} \sum_{j=1}^{k} Z_j. \tag{1.4}$$

Stellenbosch University https://scholar.sun.ac.za

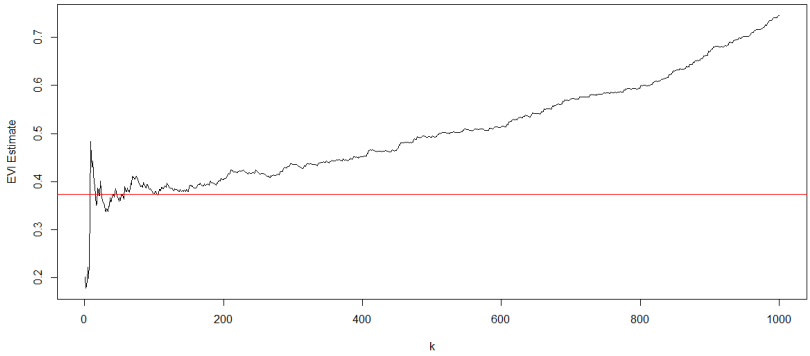

Figure 1.2: A Hill plot of the daily log-returns of Google stock from 2005 to 2019. The red line represents the Hill estimate at $k = 100$ which corresponds to Figure 1.1.

As k increases the slowly varying part ℓ_F introduces more bias into the Hill estimator. In order to account for this bias, a further second order condition is imposed on the slowly varying part ℓ_T of the tail quantile function:

$$\frac{\log \ell_T(tx) - \log \ell_T(t)}{b(t)} \quad \to \quad h_\rho(x) \quad \text{as } t \to \infty, \tag{1.5}$$

where b is some positive function that is regularly varying with index $\rho \leq 0$. This second order conditions leads to a refined exponential representation of the weighted log-spacings, as given for instance in Beirlant et al. (2004):

$$Z_j =_d \left(\gamma + b_{n,k} \left(\frac{j}{k+1}\right)^{-\rho}\right) E_j + o_p(b_{n,k}) \quad \text{for } j = 1, 2, \ldots, k \tag{1.6}$$

as $k, n \to \infty$ such that $k/n \to 0$ with $b_{n,k} = b\left(\frac{n}{k}\right)$ referred to as the bias term. Furthermore, under the second order condition it follows from de Haan and Ferreira (2006) that for $k, n \to \infty$ with $k/n \to 0$ such that $\sqrt{k}\, b_{n,k} \to M$ finite, we have that

$$\sqrt{k}\, \frac{H_{k,n} - \gamma}{\gamma} \to_d \mathcal{N}\left(\frac{M}{\gamma(1-\rho)}, 1\right).$$

In order to account for the bias of the Hill estimator, Feuerverger and Hall (1999) and Beirlant et al. (1999) developed a least squares estimator for $\gamma > 0$ based on (1.6):

Stellenbosch University https://scholar.sun.ac.za

$$\hat{\gamma}^{LS} = H_{k,n} - \frac{\bar{c}_k \sum_{j=1}^{k} (c_j - \bar{c}_k) Z_j}{\sum_{j=1}^{k} (c_j - \bar{c}_k)^2}$$

where $c_j = \left(\frac{j}{k+1}\right)^{-\hat{\rho}}$, $\bar{c}_k = \frac{1}{k}\sum_{j=1}^{k} c_j$ and $\hat{\rho}$ is a consistent estimator for ρ such as the one proposed by Fraga Alves et al. (2003b).

Caeiro et al. (2005) further assumed that b is of the form $b(x) = \gamma\beta x^{\rho}(1 + o(1))$ to yield the bias-reduced Hill estimator:

$$CH_{k,n} = \left(1 - \frac{\hat{\beta}\left(\frac{n}{k}\right)^{\hat{\rho}}}{1-\hat{\rho}}\right) H_{k,n} \tag{1.7}$$

where $\hat{\rho}$ is from Fraga Alves et al. (2003b) and $\hat{\beta}$ is from Gomes and Martins (2002). Under their respective assumptions

$$\sqrt{k}\,\frac{\hat{\gamma}-\gamma}{\gamma} \to_d \begin{cases} \mathcal{N}\left(0, \frac{(1-\rho)^2}{\rho^2}\right) & \text{for } \hat{\gamma} = \hat{\gamma}^{LS} \\ \mathcal{N}(0,1) & \text{for } \hat{\gamma} = CH_{k,n} \end{cases}$$

for $k, n \to \infty$ with $k/n \to 0$ such that $\sqrt{k}\, b_{n,k} \to M$ is finite. For a comprehensive discussion on the estimators for $\gamma > 0$ see Beirlant et al. (2004), de Haan and Ferreira (2006) and Gomes and Guillou (2015).

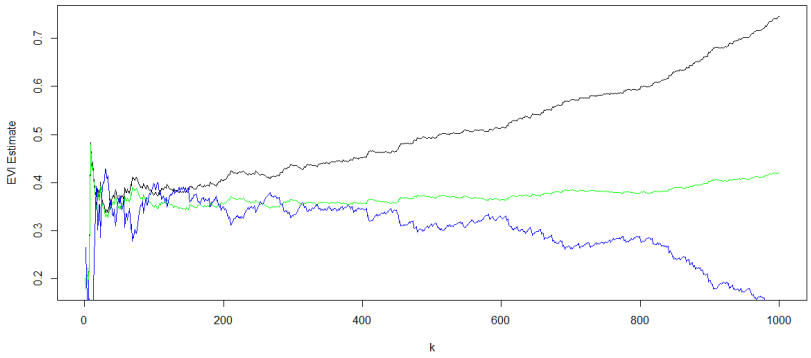

Figure 1.3: A Hill plot (black) together with the least squares (blue) and bias-corrected Hill (green) estimates for the EVI of the daily log-returns of Google stock from 2005 to 2019.

Based on the regular variation of the tail quantile function T one arrives at estimators of extreme quantiles $Q(1-p)$ with p small as proposed in Weissman

Stellenbosch University https://scholar.sun.ac.za

(1978). Indeed, under $n/k \to \infty$ one then derives that

$$\frac{Q(1-p)}{Q(1-k/n)} = \frac{T(1/p)}{T(n/k)} \sim \left(\frac{(1/p)}{(n/k)}\right)^{\gamma} = \left(\frac{k}{np}\right)^{\gamma},$$

so that with substituting $T(n/k)$ by $X_{n-k,n}$ and γ by an appropriate estimator, one arrives at

$$\hat{Q}(1-p) = X_{n-k,n}\left(\frac{k}{np}\right)^{\hat{\gamma}}.$$

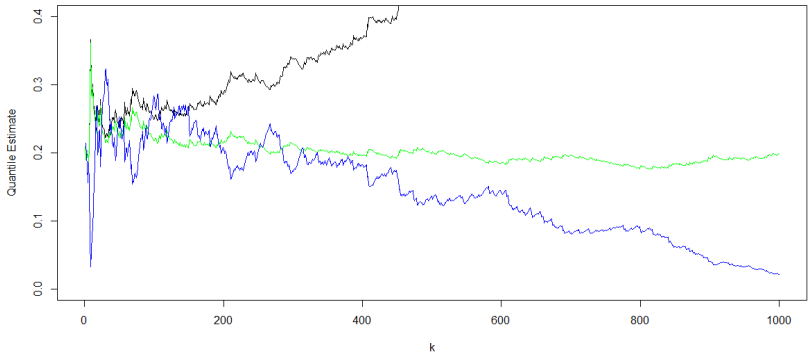

Figure 1.4: Estimates of the once-in-10-year daily log-return of Google stock using the Weissman (1978) quantile estimator together with the Hill (black), least squares (blue) and bias-corrected Hill (green) estimates for the EVI.

1.2.2 All Domains of Attraction

In the previous section we considered the case when $\gamma > 0$ where the underlying distributions are heavy-tailed. When considering a real-valued EVI $\gamma \in \mathbb{R}$, then it was established that for $\gamma > 0$ the tail decreases at a polynomial rate, for $\gamma = 0$ the tail typically decreases at an exponential rate, and for $\gamma < 0$ the tail has a finite upper bound.

Much attention has been given to developing estimators for a real-valued γ including parameter estimation of the GEV from Prescott and Walden (1980), Prescott and Walden (1983), Hosking et al. (1985) and Smith (1985), the Pickands (1975) estimator, the moment estimator from Dekkers et al. (1989), the generalized Hill estimator of Beirlant et al. (2005), and parameter estimation

Stellenbosch University https://scholar.sun.ac.za

through the use of the generalized Pareto distribution for modeling exceedances over a large threshold from Hosking and Wallis (1987) and Smith (1987). In this thesis we focus further on the generalized Hill estimator.

A bias-reduced estimator for $\gamma \in \mathbb{R}$ was first introduced by Beirlant et al. (2005), using a second order condition in the case of $\gamma \in \mathbb{R}$ as follows:

$$\lim_{t\to\infty} \frac{1}{a_1(t)} \left(\frac{T(xt) - T(t)}{a(t)} - h_\gamma(u) \right) = \frac{h_{\gamma+\tilde{\rho}}(u) - h_\gamma(u)}{\tilde{\rho}} \tag{1.8}$$

for some positive function a_1 that is regularly varying with index $\rho_1 < 0$. A specific case in which this condition holds is when

$$T(x) = \begin{cases} Cx^\gamma \left(1 + Dx^{\rho_1}(1 + o(1))\right) & \text{if } \gamma > 0, \\ T(\infty) - Cx^\gamma \left(1 + Dx^{\rho_1}(1 + o(1))\right) & \text{if } \gamma < 0, \end{cases} \tag{1.9}$$

as $x \to \infty$, for some $C > 0$, $D \in \mathbb{R}$, and $\rho_1 < 0$. In the case $\gamma = 0$ all relevant examples have the second order parameter ρ_1 equal to 0 leading to substantial bias in such cases. In the case $\gamma > 0$ we have that $\rho_1 = \rho$. Defining the generalized log-spacings Y_j as

$$Y_j = (j+1) \left(\log \frac{X_{n-j} H_{j,n}}{X_{n-j-1,n} H_{j+1,n}} - \log\left(1 + \frac{1}{j}\right) + \frac{1}{j} \right), \tag{1.10}$$

Beirlant et al. (2005) showed that under (1.9) it follows that:

$$Y_j = \gamma + b_{n,k} \left(\frac{j}{k+1} \right)^{-\tilde{\rho}} + \epsilon_j \qquad \text{for } j = 1, 2, \ldots, k, \tag{1.11}$$

where $b(x) = \gamma\tilde{\beta}x^{\tilde{\rho}}(1 + o(1))$ as $x \to \infty$, $\tilde{\rho}$ is a function of ρ_1, and ϵ_j are independent error terms with asymptotic mean 0 as $k, n \to \infty$ and $k/n \to 0$. This is a generalization of the result (1.6) to real-valued γ.

Based on (1.11) the following estimators were then proposed in Beirlant et al. (2005): the generalized Hill estimator for $\gamma \in \mathbb{R}$

$$GH_{k,n} = \frac{1}{k} \sum_{j=1}^{k} Y_j,$$

and the generalized least squares estimator

$$\hat{\gamma}^{LS} = GH_{k,n} - \frac{\bar{d}_k \sum_{j=1}^{k} \left(d_j - \bar{d}_k\right) Y_j}{\sum_{j=1}^{k} \left(d_j - \bar{d}_k\right)^2},$$

Stellenbosch University https://scholar.sun.ac.za

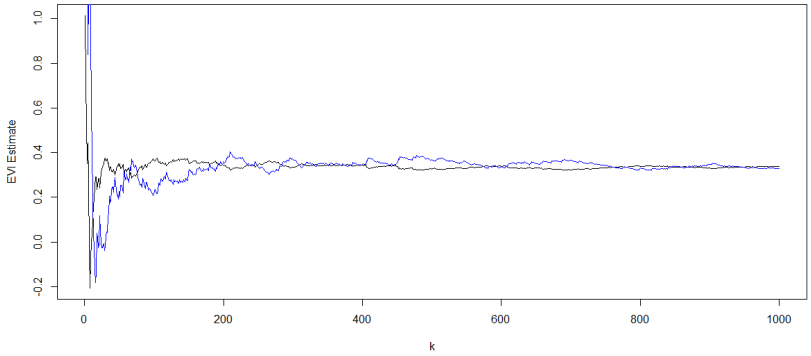

Figure 1.5: The generalized Hill (black) and least squares (blue) estimates for the EVI of the daily log-returns of Google stock from 2005 to 2019.

where $d_j = \left(\frac{j}{k+1}\right)^{-\tilde{\rho}}$ and $\bar{d}_k = \frac{1}{k}\sum_{j=1}^{k} d_j$.

Furthermore for $k, n \to \infty$ with $k/n \to 0$ such that $\sqrt{k}\, b_{n,k} \to M$ is finite one has

$$\sqrt{k}\,\frac{\hat{\gamma}-\gamma}{\gamma} \to_d \begin{cases} \mathcal{N}\left(\frac{M}{\gamma\,(1-\rho)}, \sigma_1^2\right) & \text{for } \hat{\gamma} = GH_{k,n} \\ \mathcal{N}\left(0, \sigma_2^2\right) & \text{for } \hat{\gamma} = \hat{\gamma}^{LS} \end{cases}$$

where σ_1^2 and σ_2^2 will be discussed further in Chapter 3 below.

1.2.3 Multivariate Extremes

In the previous sections we focused on the statistical analysis of univariate extremes. However, in many statistical problems the random variables are multivariate and the extremal properties of their joint statistical behavior is of interest. The principle challenge with the analysis of multivariate extremes is that multivariate observations do not have a canonical ordering, which in turn means that the definition of a multivariate quantile does not readily follow as in the univariate case.

There are a range of approaches to modeling multivariate extremes, some of which transform the multivariate problem to a univariate one. Based on the assumption that the underlying distribution is multivariate regularly varying, any convex combination of the random vector has a univariate regularly varying tail with the same EVI, which effectively transforms the multivariate problem

Stellenbosch University https://scholar.sun.ac.za

to a univariate one. Kim and Lee (2017) use the above-mentioned property of a multivariate regularly varying distribution to propose an estimator for the EVI that is a weighted average of Hill (1975) estimators for every convex combination of the random vector, in a manner which minimizes the variability of the estimator.

The recent development of empirical center-outward quantiles in $\mathbb{R}^d$ by Hallin (2017) provides a new mathematical basis to model multivariate extremes. The empirical center-outward quantiles of Hallin (2017) result from an optimal transport of the unit ball $\mathbb{S}^d$ to $\mathbb{R}^d$ based on theory developed by Rockaffelar (1996). This generalization of the notion of a quantile to d-dimensional space leads to applications to multivariate risk measurement, including the estimation of the EVI for a multivariate sample.

1.3 Thesis Outline

This thesis consists of four chapters. The present chapter is introductory and provides background and context regarding the role of extreme value analysis in understanding the statistical behavior of extreme events. We here provided an overview of the mathematical theory underlying extreme value analysis, which includes the assumptions and results on which the statistical inference of extreme quantiles is developed.

In the second chapter, ridge regression estimators for the EVI are proposed for both Pareto-type distributions ($\gamma > 0$) and all domains of attraction ($\gamma \in \mathbb{R}$) starting from the regression models (1.6) and (1.11). The proposed ridge regression estimator for $\gamma > 0$ is effectively a combination of the Hill (1975) and least-squares estimators, and utilizes their bias-variance trade-off as a function of k to minimize its asymptotic mean squared error. Similar results hold for the ridge regression estimators for $\gamma \in \mathbb{R}$. These bias-reduced estimators are applied to daily wind speed data to illustrate their accuracy and utility.

A novel pivotal statistic is proposed in the third chapter with which normal and saddlepoint confidence intervals are developed for extreme Pareto-type quantiles. To this end we consider the extreme quantile estimator of Weissman (1978) and its bias-reduced extension of Matthys et al. (2004) together with the Hill estimator and its bias-corrected counterpart, and the ridge regression estimators discussed in the second chapter. The utility of the resulting confidence intervals are illustrated by means of simulation and practical case studies on reinsurance

Stellenbosch University https://scholar.sun.ac.za

data.

The fourth chapter discusses the application of empirical center-outward quantiles to multivariate risk measurement. Hallin (2017) defined the empirical center-outward quantiles in $\mathbb{R}^d$ as the inverse of the empirical center-outward distribution function in $\mathbb{R}^d$, which is an optimal transportation from $\mathbb{R}^d$ to the unit ball $\mathbb{S}^d$ that satisfies a Glivenko-Cantelli theorem. In this chapter, a continuous version of the empirical center-outward quantile function of Hallin (2017) is proposed. Furthermore, risk measures based on the optimal transportation are developed. Finally, a new approach to estimate the EVI for a multivariate regularly varying distribution is presented. The proposed continuous center-outward quantiles and risk measures are applied to the daily log-returns of Google and Apple share prices, as well as to Danish fire insurance data in order to illustrate their value.

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Chapter 2

Ridge Regression Estimators for the Extreme Value Index

We consider bias reduced estimators of the extreme value index (EVI) in case of Pareto-type distributions and under all max-domains of attraction. To this purpose we revisit the regression approach started in Feuerverger and Hall (1999) and Beirlant et al. (1999) in the case of a positive EVI, and in Beirlant et al. (2005) for real-valued EVI. We generalize these approaches using ridge regression exploiting the mathematical fact that the bias tends to 0 when the number of top data points used in the estimation is decreased. The penalty parameter is selected by minimizing the asymptotic mean squared error of the proposed estimator. The accuracy and utility of the ridge regression estimators are studied using simulations and are illustrated with case studies on reinsurance claim size data as well as daily wind speed data.

2.1 Introduction

Extreme value methodology receives growing attention in order to model the occurrence of rare events with high impact in various fields of application such as finance, insurance, hydrology and climatology. Estimation of the extreme value index (EVI) γ is then a crucial topic in extreme value methodology, assuming that the underlying distribution satisfies the max-domain of attraction condition, i.e. assuming that the maximum of independent and identically distributed observations $X_1, X_2, \ldots, X_n$ can be approximated by the generalized extreme

Stellenbosch University https://scholar.sun.ac.za

value distribution:

$$\mathbb{P}\left(a_n^{-1}\left(\max_{i=1,\ldots,n} X_i - b_n\right) \leq y\right) \underset{n\to\infty}{\longrightarrow} G_\gamma(y) = \exp\left(-(1+\gamma y)^{-1/\gamma}\right) \quad (2.1)$$

for $1+\gamma\, y > 0$ where $b_n \in \mathbb{R}$, $a_n > 0$ and $\gamma \in \mathbb{R}$ are the location, scale and shape parameters, respectively. The EVI γ is a measure of the tail-heaviness of the distribution of X with a larger value of γ implying a heavier tail of F.

In the specific case of $\gamma > 0$ equation (2.1) corresponds to the set of Pareto-type distributions with right tail function (RTF) given by

$$\bar{F}(x) = 1 - F(x) = P(X > x) = x^{-\frac{1}{\gamma}}\, \ell(x) \quad (2.2)$$

where ℓ is a slowly varying function at infinity, i.e.

$$\frac{\ell(ty)}{\ell(t)} \underset{t\to\infty}{\longrightarrow} 1 \quad \text{for every } y > 1. \quad (2.3)$$

The estimation of γ under (2.2) has received most attention starting with the Hill (1975) estimator

$$H_{k,n} = \frac{1}{k}\sum_{j=1}^{k} \log X_{n-j+1,n} - \log X_{n-k,n} = \frac{1}{k}\sum_{j=1}^{k} Z_j,$$

where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ denote the ordered observations and where Z_j $(j = 1, \ldots, n)$ denote the weighted log-spacings

$$Z_j = j\, \log \frac{X_{n-j+1,n}}{X_{n-j,n}}. \quad (2.4)$$

While this estimator has minimal variance, its bias however increases with a growing number k of top data used in the estimator. The factor $\ell(x)$ describes the deviation from the simple Pareto model with RTF $x^{-1/\gamma}$. The more ℓ differs from a constant, i.e. the slower the convergence in (2.3), the larger the bias of the available estimators. Much research has been performed in order to estimate the value of k which minimizes the mean squared error (for an overview see for instance Matthys and Beirlant (2000)) or to construct estimators which exhibit reduced bias starting with the papers of Feuerverger and Hall (1999) and Beirlant et al. (1999). Both these papers proposed to use regression models on Z_j using the scaled ranks $\left(\frac{j}{k+1}\right)^{-\rho}$ as a covariate and with the extreme value index γ appearing as the intercept. The parameter ρ is specified by the second-order condition from Hall (1982) which is given in section 2.2.1.

Stellenbosch University https://scholar.sun.ac.za

In this paper we concentrate on bias reduction with the aim of producing estimators that are much more horizontal over k than other popular EVI estimators discussed in this section, hence facilitating application in practice to a great extent. The main drawback of bias reduced EVI estimators is that they naturally exhibit larger variance, especially for small k values. In a pioneering paper, Caeiro et al. (2005) proposed a corrected Hill estimator $CH_{k,n}$ given in (1.7) with the same asymptotic variance as $H_{k,n}$ and excellent bias and MSE characteristics while specifying the convergence in (2.3) up to a second order. Cai et al. (2013) provided a reduced bias estimator in case γ is close to 0 using a probability weighted moment approach. As an alternative, we propose to reduce the increase in variance by penalizing the bias reduction for small values of k. Beirlant et al. (2019) penalized the use of a bias reducing extension of the simple Pareto model with RTF $x^{-1/\gamma}$ when modeling the exceedances X/t given $X > t$ for large values of t. Here we use ridge regression on the linear regression model on the Z_j. This reduces the bias reduction effect for smaller k, which leads to bias reduced estimators following the behavior of the classical Hill estimator with small bias and minimal variance at small k. For a general review on ridge regression, we refer to e.g. Hastie et al. (2008).

Under the general max-domain of attraction condition (2.1) bias-reduced estimation of $\gamma \in \mathbb{R}$ has received almost no attention. To our knowledge bias reduced versions of the popular moment estimator (see Dekkers et al. (1989)) and of the estimators based on the peaks over threshold (POT) approach fitting the generalized Pareto distribution to the distribution of $X - t$ given $X > t$ for large thresholds t (see Smith (1987) and Hosking and Wallis (1987)) are not available. In Beirlant et al. (2005) a generalization $GH_{k,n}$ of the Hill estimator and a non-linear regression model for the generalized weighted log-spacings Y_j $(j = 1, \ldots, n)$ using the covariates $\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}$ was proposed as follows:

$$Y_j = (j+1)\left(\log \frac{X_{n-j}H_{j,n}}{X_{n-j-1,n}H_{j+1,n}} - \log\left(1+\frac{1}{j}\right) + \frac{1}{j}\right), \tag{2.5}$$

and

$$GH_{k,n} = \frac{1}{k}\sum_{j=1}^{k} Y_j. \tag{2.6}$$

The parameter $\tilde{\rho}$ is deduced from the generalized second-order condition which is given in section 2.2. For a comparison of $GH_{k,n}$ with the maximum likelihood estimator based on generalized Pareto modeling for exceedances over a high threshold and with the moment estimator (see Dekkers et al. (1989)) we can refer to Beirlant et al. (2005). In Beirlant et al. (2005) the focus was on optimal selection of k, while the properties of the corresponding least squares regression

Stellenbosch University https://scholar.sun.ac.za

estimator of the intercept were not studied. We next propose ridge regression of Y_j on $\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}$ $(j = 1, \ldots, k)$ in order to obtain bias reduced estimators of $\gamma \in \mathbb{R}$ with good variance properties for small k.

The penalty parameter in the ridge regression will be chosen on the basis of asymptotic mean squared error arguments. The regression models and the ridge regression will be introduced in section 2.2, both for the case of Pareto-type distributions with $\gamma > 0$ and in general with $\gamma \in \mathbb{R}$. A numerical study and some practical illustrations of the proposed method will be proposed in sections 2.3 and 2.4. The technical details concerning the asymptotic mean squared errors of the ridge regression estimators are deferred to Addendum A in the Appendix.

2.2 Ridge regression estimators

In this section we recollect the original regression models for Z_j and Y_j and introduce the ridge regression approach. We also discuss the selection of the penalty parameter. Assume that the RTF is continuous and strictly decreasing and denote the inverse function of the RTF $\bar{F}$ by $\bar{F}^{-1}$, the max-domain of attraction condition is introduced in terms of the tail quantile function

$$T(x) = \bar{F}^{-1}(x^{-1}) = F^{-1}\left(1 - \frac{1}{x}\right) \quad \text{for } x > 1,$$

which denotes the outcome level of X which is surpassed with probability $1/x$. We first consider the case $\gamma > 0$, after which the generalization to all max-domains will be discussed.

2.2.1 Pareto-type distributions

Assume that the RTF satisfies (2.2). The equivalent condition based on the tail quantile function T is then given by

$$T(x) = x^{\gamma}\ell_T(x), \tag{2.7}$$

where ℓ_T is a slowly varying function at infinity, i.e. ℓ_T also satisfies the condition (2.3). As is usual when studying the estimation of $\gamma > 0$, it is assumed that ℓ_T follows a second-order condition introduced in Hall (1982): there exists some

Stellenbosch University https://scholar.sun.ac.za

positive function b that is regularly varying with index $\rho < 0$, i.e. $b(x) = x^{\rho}\ell_b(x)$ for some slowly varying ℓ_b, such that for all $x > 0$:

$$\frac{\log \ell_T(tx) - \log \ell_T(t)}{b(t)} \underset{t\to\infty}{\longrightarrow} \frac{x^{\rho} - 1}{\rho}. \tag{2.8}$$

Under (2.8), Beirlant et al. (2002) showed that the weighted log-spacings Z_j can be represented using the following non-linear regression model:

$$Z_j = \gamma + b_{n,k}\, c_j + \epsilon_j \qquad \text{for } j = 1, 2, \ldots, k, \tag{2.9}$$

with slope $b_{n,k} = b\left(\frac{n}{k}\right)$, covariate values $c_j = \left(\frac{j}{k+1}\right)^{-\rho}$, and where ϵ_j are independent error terms with an asymptotic mean of 0 and variance of γ^2 as $k, n \to \infty$ and $k/n \to 0$. The specification of the error terms ϵ_j can be found in section 4.4 in Beirlant et al. (2004).

Ridge regression, originally developed by Hoerl and Kennard (1970) to solve multi-collinearity problems among predictors, can be viewed as the solution of the minimization of the ridge loss function, which adds an L^2 penalty term to the classical regression sum of squares. Here the ridge regression loss function is given by

$$L_k\left(\gamma, b_{n,k}; \tau\right) = \sum_{j=1}^{k} \left(Z_j - \gamma - b_{n,k}\, c_{j,k}\right)^2 + \tau k b_{n,k}^2,$$

the minimisation of which with respect to γ and $b_{n,k}$ leads to the ridge regression estimators

$$\hat{b}^{+}_{n,k}(\tau) = \frac{S_{cZ}}{S_{cc} + \tau} \qquad \text{and} \qquad \hat{\gamma}^{+}_{k}(\tau) = \bar{Z}_k - \bar{c}\,\hat{b}^{+}_{n,k}(\tau), \tag{2.10}$$

where $S_{cZ} = \frac{1}{k}\sum_{j=1}^{k}\left(c_j - \bar{c}\right) Z_j$, $S_{cc} = \frac{1}{k}\sum_{j=1}^{k}\left(c_j - \bar{c}\right)^2$, and $\bar{c} = \frac{1}{k}\sum_{j=1}^{k} c_j$. Note that for $\tau = 0$, the ridge regression estimator $\hat{\gamma}^{+}_{k}(0)$ equals the least-squares regression estimator $\hat{\gamma}^{\mathrm{LS}+}_{k}$, and for $\tau \to \infty$, $\hat{\gamma}_k(\tau)$ converges to the Hill estimator $H_{k,n} = \bar{Z}_k$ in (1.4).

The loss function L_k balances the sum of squares and the penalty term on the bias. The effect of the penalty term is to shrink $b_{n,k}$ to zero by increasing the penalty parameter τ. The larger τ, the larger the contribution of the penalty term to the loss function, and the stronger the tendency to shrink the regression slope coefficient $b_{n,k}$, or equivalently, to gradually move from the

Stellenbosch University https://scholar.sun.ac.za

least squares estimator to the Hill estimator. In fact, $b_{n,k} = b(n/k)$ tends to 0 with increasing n/k, and hence for the smallest values of k the bias of the Hill estimator is smallest and bias correction is not needed, and hence the penalty on the bias coefficient $b_{n,k}$ should be increased. We therefore allow τ to depend on k.

We choose the penalty parameter $\tau = \tau_k^+$ to minimize the asymptotic mean squared error (AMSE) of the ridge regression estimator $\hat{\gamma}_k^+(\tau)$. It follows from (2.9) and (2.10) that the ridge regression estimator can be written as

$$\hat{\gamma}_k^+(\tau) = \bar{Z}_k - \frac{\bar{c}\, S_{cZ}}{S_{cc}+\tau} = \frac{1}{k}\sum_{j=1}^{k} \lambda_j(\tau)\, Z_j = \frac{1}{k}\sum_{j=1}^{k} \lambda_j(\tau)\ (\gamma + b_{n,k}\, c_j + \epsilon_j)$$

where $\lambda_j(\tau) = 1 - \frac{\bar{c}\,(c_j - \bar{c})}{S_{cc}+\tau}$, and

$$AMSE\left(\hat{\gamma}_k^+(\tau)\right) = \frac{1}{k^2}\sum_{j=1}^{k} \left(\lambda_j(\tau)\right)^2 Var\left(\epsilon_j\right) + \left(\frac{1}{k}\sum_{j=1}^{k} \lambda_j(\tau)\left(\gamma + b_{n,k}\, c_j\right) - \gamma\right)^2$$

$$= \frac{\gamma^2}{k}\left(1 + \frac{\bar{c}^2\, S_{cc}}{\left(S_{cc}+\tau\right)^2}\right) + \bar{c}^2\, b_{n,k}^2 \left(1 - \frac{S_{cc}}{S_{cc}+\tau}\right)^2.$$

Setting the derivative of the AMSE with respect to τ equal to 0, it follows that the optimal τ value is a solution of the equation

$$\frac{k\, b_{n,k}^2\, \tau - \gamma^2}{\left(S_{cc}+\tau\right)^3} = 0.$$

Excluding the solution $\tau \to \infty$, we arrive at

$$\tau_k^+ = \frac{\gamma^2}{k\, b_{n,k}^2}.$$

This solution τ_k^+ minimizes the AMSE for all k and is always positive. It can therefore be considered as the asymptotic optimal choice for τ. We propose to estimate τ_k^+ assuming that the slowly varying function ℓ_b in (2.8) is basically a constant:

$$b(x) = \gamma\beta x^{\rho}(1 + o(1)) \text{ as } x \to \infty \tag{2.11}$$

for some $\beta \in \mathbb{R}$. We then propose to use the estimator $\hat{\beta}_k$ from Gomes and Martins (2002)

$$\hat{\beta}_k = \left(\frac{k}{n}\right)^{\hat{\rho}} \frac{\left(\sum_{j=1}^{k} c_{j,k^\star}\right)\left(\sum_{j=1}^{k} Z_j\right) - k\sum_{j=1}^{k} c_{j,k} Z_j}{\left(\sum_{j=1}^{k} c_{j,k}\right)\left(\sum_{j=1}^{k^\star} c_{j,k} Z_j\right) - k\sum_{j=1}^{k} c_{j,k^\star}^2 Z_j}$$

Stellenbosch University https://scholar.sun.ac.za

to obtain

$$\hat{\tau}_k^+ = \frac{1}{k\,\hat{\beta}_k^2\,\left(\frac{n}{k}\right)^{2\hat{\rho}}}. \tag{2.12}$$

Consistent estimators $\hat{\rho}$ for the second-order parameter ρ have been proposed for instance in Fraga Alves et al. (2003a). Alternatively, in order to obtain plots of estimators $\hat{\gamma}_k$ which are as stable as possible as a function of k, one can choose ρ to minimize the variance over a large range of k values using the simple least squares estimator for γ based on (2.9):

$$\hat{\rho}_n = \arg \min_{-2<\tilde{\rho}<-\frac{1}{2}} \sum_{k=3}^{k^*} \left(\hat{\gamma}_k^{\mathrm{LS}+} - \bar{\hat{\gamma}}^{\mathrm{LS}+}\right)^2$$

where $\hat{\gamma}_k^{\mathrm{LS}+} = \bar{Z}_k - \frac{\bar{c}\,\sum_{j=1}^{k}(c_j-\bar{c})\,Z_j}{\sum_{j=1}^{k}(c_j-\bar{c})^2}$ and $\bar{\hat{\gamma}}^{\mathrm{LS}+}$ denotes the average of the estimates $\hat{\gamma}_k^{\mathrm{LS}+}$ for $k = 1, 2, , \ldots, k^*$ and $k^* = \left[n^{.99}\right]$. This range is chosen to limit the influence of the largest and smallest data points on the specification of $\hat{\rho}_n$. In the simulations and case studies we use the minimum variance option, while the differences between the use of the Fraga Alves et al. (2003a) estimator and the present choice is rather limited.

The resulting ridge regression estimator for $\gamma > 0$ is therefore given by

$$\hat{\gamma}_k^+ = \hat{\gamma}_k^+\left(\hat{\tau}_k^+\right). \tag{2.13}$$

Remark 2.1. The parameter $1/\tau = \frac{k\,b_{n,k}^2}{\gamma^2}$ was used in Guillou and Hall (2001) to propose an adaptive choice of k when using the Hill estimator under (2.8) and (2.11). Using the Hill estimator for γ and the least squares estimator for $b_{n,k}$ they proposed to use $H_{k,n}$ at the smallest value k for which

$$\sqrt{\frac{k}{12}}\,\frac{|\hat{b}_{n,k}^{\mathrm{LS}+}|}{H_{k,n}} > 1.25,$$

so that for values of k larger than this adaptive choice the use of the Hill estimator is in some sense "rejected", suggesting then to use a bias reduced estimator. Here, rather than choosing one particular k, the ridge regression approach with the penalty parameter τ depending on k, leads to an automatic combination of the Hill and the least squares estimator. See Figure 2.10 for a plot of $1/\hat{\tau}_k^+$ in a case study. □

Stellenbosch University https://scholar.sun.ac.za

Remark 2.2. In statistical learning theory, the penalty parameter τ is often chosen to minimize the prediction error of the ridge regression fit. Cross-validation is considered to be the simplest and most widely used method of estimating the prediction error; see for instance Chapter 7 section 10 in Hastie et al. (2008). In m-fold cross-validation the data is divided into m equally-sized groups and each group is taken in turn as the validation set, while the other groups are combined to train the model. The m corresponding prediction errors, here denoted as $CV_1, CV_2, \ldots, CV_m$, are then averaged to give the m-fold cross-validation error. This gives

$$CV(\tau; \underline{Z}_k) = \frac{1}{m} \sum_{i=1}^{m} CV_i(\tau; \underline{Z}_k)$$

where

$$CV_i(\tau; \underline{Z}_k) = \frac{1}{|K_i|} \sum_{j \in K_i} \left(Z_j - \frac{1}{|K_i^c|} \sum_{l \in K_i^c} Z_l - \hat{b}_{n,K_i} \left(c_j - \frac{1}{|K_i^c|} \sum_{l \in K_i^c} c_l \right) \right)^2,$$

and $\hat{b}_{n,K_i} = \frac{\sum_{l \in K_i^c} (c_l - \bar{c}_i)\, Z_l}{\sum_{l \in K_i^c} (c_l - \bar{c}_i)^2 + |K_i^c|\, \tau}$ where the sets of indices $K_1, K_2, \ldots, K_m$ are equal-sized partitions of $\{1, 2, \ldots, k\}$, with the complement of the set defined as $K_i^c = \{1, 2, \ldots, k\} \setminus K_i$, and $\bar{c}_i = \frac{1}{|K_i^c|} \sum_{l \in K_i^c} c_l$. $CV_{i,k}$ is the prediction error of the ridge regression model trained on the data indexed by K_i^c and tested on the data indexed by K_i. The optimal penalty parameter is then estimated as

$$\arg\min_{\tau} CV(\tau; \underline{Z}_k).$$

From the simulation study the corresponding ridge regression estimator turned out to perform less good than those obtained from using (2.12). We do not report here the estimates resulting from using cross validation. □

We end the discussion for the case of Pareto-type distributions stating the asymptotic distribution of the ridge regression estimator $\hat{\gamma}_k^+$ defined in (2.13). The derivation of this result is discussed in Addendum A in the Appendix.

Theorem 2.1. *Assume* (2.8) *and* (2.11) *hold. Assume* ρ *is estimated by a consistent estimator* $\hat{\rho}$. *Then, as* $k, n \to \infty$ *and* $k/n \to 0$ *and* $\sqrt{k} b_{n,k} \to M \in \mathbb{R}$,

$$\sqrt{k} \left(\hat{\gamma}_k^+ - \gamma \right) \to_d \mathcal{N}\left(\mu_M, \sigma_M^2 \right).$$

where $\mu_M = \frac{M\gamma^2(1-\rho)(1-2\rho)}{\rho^2 M^2 + \gamma^2 (1-\rho)^2 (1-2\rho)}$ *and* $\sigma_M = \gamma \sqrt{1 + \frac{M^4 \rho^2 (1-2\rho)}{\{M^2\rho^2 + \gamma^2(1-\rho)^2(1-2\rho)\}^2}}$.

Stellenbosch University https://scholar.sun.ac.za

Note that for $M \to 0$ the asymptotic bias and variance are those of the Hill estimator, i.e. 0 and γ^2. For large values of k, expressed here by $M \to \infty$ the asymptotic bias and variance are those of the least squares estimator $\hat{\gamma}_k^{\mathrm{LS}+}$, i.e. 0 and $\gamma^2((1-\rho)/\rho)^2$. This confirms in an asymptotic way the above assertions that the ridge regression estimator constitutes a compromise between the Hill and least squares estimator.

2.2.2 All max-domains of attraction

Assume that the distribution of X satisfies (2.1) for some $\gamma \in \mathbb{R}$. The equivalent (first-order) condition based on the tail quantile function T is that

$$\lim_{t\to\infty} \frac{T(xt) - T(t)}{a(t)} = h_\gamma(u) := \int_1^x u^{\gamma-1} du.$$

One can further consider a second order condition:

$$\lim_{t\to\infty} \frac{1}{a_1(t)} \left(\frac{T(xt) - T(t)}{a(t)} - h_\gamma(u) \right) = \frac{h_{\gamma+\tilde{\rho}}(u) - h_\gamma(u)}{\tilde{\rho}} \tag{2.14}$$

for some positive function a_1 that is regularly varying with index $\rho_1 < 0$. This condition is for instance fulfilled when

$$T(x) = \begin{cases} Cx^\gamma \left(1 + Dx^{\rho_1}(1+o(1))\right) & \text{if } \gamma > 0, \\ T(\infty) - Cx^\gamma \left(1 + Dx^{\rho_1}(1+o(1))\right) & \text{if } \gamma < 0, \end{cases} \tag{2.15}$$

as $x \to \infty$, for some $C > 0$, $D \in \mathbb{R}$, and $\rho_1 < 0$. In case $\gamma > 0$ we have that $\rho_1 = \rho$. Under (2.15), Beirlant et al. (2005) showed that the generalized weighted log-spacings Y_j can be represented using the following linear regression model which is basically identical to the model (2.9):

$$Y_j = \gamma + b_{n,k}\, c_j + \epsilon_j \qquad \text{for } j = 1, 2, \ldots, k, \tag{2.16}$$

where $b(x) = \gamma\tilde{\beta}x^{\tilde{\rho}}(1+o(1))$ $(x \to \infty)$ with $\tilde{\rho}$ deduced from ρ_1 as explained in section 2 in Beirlant et al. (2005). Here $c_j = \left(\frac{j}{k+1}\right)^{-\tilde{\rho}}$ and ϵ_j are independent error terms with asymptotic mean 0 as $k, n \to \infty$ and $k/n \to 0$. Ridge regression applied to the linear regression model (2.16) then yields (similar to (2.10)):

$$\hat{b}_{n,k}(\tau) = \frac{S_{cY}}{S_{cc} + \tau} \qquad \text{and} \qquad \hat{\gamma}_k(\tau) = \bar{Y}_k - \bar{c}\,\hat{b}_{n,k}(\tau) \tag{2.17}$$

where $S_{cY} = \frac{1}{k}\sum_{j=1}^k (c_j - \bar{c})\, Y_j$. Here the penalty parameter $\tau > 0$ allows for compromises between the generalized Hill estimator $GH_{k,n}$ when $\tau \to \infty$ and

Stellenbosch University https://scholar.sun.ac.za

the least squares regression estimator $\hat{\gamma}_k^{\mathrm{LS}} = \hat{\gamma}_k(0)$.

We again choose the penalty parameter $\tau = \tau_k$ to minimize the asymptotic mean squared error of the ridge regression estimator $\hat{\gamma}_k(\tau)$. The AMSE is derived in Addendum B using (2.16) and (2.17) and is given by

$$\begin{aligned} AMSE\left(\hat{\gamma}_k(\tau)\right) =& \frac{(1-\gamma)^2}{k}\left(1+\frac{\bar{c}^2 S_{cc}}{(S_{cc}+\tau)^2}\right) \\ &+ \frac{2\gamma}{k}\left(\xi_0(\gamma)+\bar{c}^2 S_{cc}\left(\frac{\xi_1(\gamma)}{S_{cc}+\tau}+\frac{\xi_2(\gamma)}{(S_{cc}+\tau)^2}\right)\right) \\ &+ \bar{c}^2\, b_{n,k}^2\left(1-\frac{S_{cc}}{S_{cc}+\tau}\right)^2, \end{aligned}$$

where

$$\xi_0(\gamma) = \begin{cases} 1 & \text{if } \gamma \geq 0 \\ \frac{2(1-\gamma)}{1-2\gamma} & \text{if } \gamma < 0, \end{cases}$$

$$\xi_1(\gamma) = \begin{cases} \frac{-\tilde{\rho}}{(1-\tilde{\rho})S_{cc}} & \text{if } \gamma \geq 0 \\ \frac{2(1-\gamma)}{1-2\gamma}\,\frac{-\tilde{\rho}}{(1-\gamma-\tilde{\rho})\,S_{cc}} & \text{if } \gamma < 0, \end{cases}$$

$$\xi_2(\gamma) = \begin{cases} \frac{1}{1-\tilde{\rho}} & \text{if } \gamma \geq 0 \\ \frac{2(1-\gamma)}{1-2\gamma}\,\frac{1-\gamma}{1-\gamma-\tilde{\rho}} & \text{if } \gamma < 0. \end{cases}$$

By setting the derivative of the AMSE with respect to τ equal to 0, the optimal $\tau = \tau_k$ is a solution of the equation

$$\frac{\left(k\, b_{n,k}^2 - \gamma\xi_1(\gamma)\right)\tau - (1-\gamma)^2 - \gamma\left(2\xi_2(\gamma) + S_{cc}\,\xi_1(\gamma)\right)}{(S_{cc}+\tau)^3} = 0.$$

The finite solution equals

$$\tau_k = \frac{(1-\gamma)^2 + \gamma\left(2\xi_2(\gamma) + S_{cc}\,\xi_1(\gamma)\right)}{k\, b_{n,k}^2 - \gamma\xi_1(\gamma)},$$

which minimizes the AMSE error for all k. However, it is required that τ_k is positive, and therefore if the solution τ_k is negative, $\tau = 0$ should be used. We propose the following estimator for τ_k:

$$\hat{\tau}_k = \frac{\left(1-\hat{\gamma}_k^{\mathrm{LS}}\right)^2 + \hat{\gamma}_k^{\mathrm{LS}}\left(2\xi_2\left(\hat{\gamma}_k^{\mathrm{LS}}\right) + S_{cc}\,\xi_1\left(\hat{\gamma}_k^{\mathrm{LS}}\right)\right)}{k\left(\hat{\gamma}_k^{\mathrm{LS}}\hat{\bar{\beta}}_k\left(\frac{n}{k}\right)^{\hat{\tilde{\rho}}}\right)^2 - \hat{\gamma}_k^{\mathrm{LS}}\xi_1\left(\hat{\gamma}_k^{\mathrm{LS}}\right)}, \tag{2.18}$$

Stellenbosch University https://scholar.sun.ac.za

where $\hat{\gamma}_k^{\mathrm{LS}} = \hat{\gamma}_k(0) = \bar{Y}_k - \frac{\bar{c}\sum_{j=1}^{k}(c_j-\bar{c})\,Y_j}{\sum_{j=1}^{k}(c_j-\bar{c})^2}$ is the least squares estimator, and $\hat{\tilde{\rho}}$ and $\hat{\tilde{\beta}}$ are estimators of $\tilde{\rho}$ and $\tilde{\beta}$. As in the case of Pareto-type distributions we propose to estimate $\tilde{\beta}$ by

$$\hat{\tilde{\beta}}_k = \left(\frac{k}{n}\right)^{\hat{\tilde{\rho}}} \frac{\left(\sum_{j=1}^{k} c_{j,k}\right)\left(\sum_{j=1}^{k} Y_j\right) - k\sum_{j=1}^{k} c_{j,k}Y_j}{\left(\sum_{j=1}^{k} c_{j,k}\right)\left(\sum_{j=1}^{k} c_{j,k}Y_j\right) - k\sum_{j=1}^{k} c_{j,k}^2 Y_j}.$$

Fraga Alves et al. (2003a) proposed consistent estimators $\hat{\rho}_1$ of ρ_1. Alternatively we here also propose to use

$$\hat{\tilde{\rho}} = \arg \min_{-2<\tilde{\rho}<-\frac{1}{2}} \sum_{k=3}^{k^*} \left(\hat{\gamma}_k^{\mathrm{LS}} - \bar{\hat{\gamma}}^{\mathrm{LS}}\right)^2.$$

The generalized ridge regression estimator for $\gamma \in \mathbb{R}$ is therefore given by

$$\hat{\gamma}_k = \begin{cases} \hat{\gamma}_k\left(\hat{\tau}_k\right) & \text{if } \hat{\tau}_k \geq 0 \\ \hat{\gamma}_k(0) & \text{if } \hat{\tau}_k < 0. \end{cases} \tag{2.19}$$

Similarly as in Theorem 1, we state the asymptotic distribution of $\hat{\gamma}_k$. To this end we set

$$\tau_M = \frac{(1-\gamma)^2 + \gamma\{2\xi_2(\gamma) + \frac{\tilde{\rho}^2}{(1-\tilde{\rho})^2(1-2\tilde{\rho})}\xi_1(\gamma)\}}{M^2 - \gamma\xi_1(\gamma)}.$$

Theorem 2.2. *Assume* (2.15) *holds. Assume* $\tilde{\rho}$ *is estimated by a consistent estimator* $\hat{\tilde{\rho}}$. *Then, as* $k, n \to \infty$ *and* $k/n \to 0$ *and* $\sqrt{k}b_{n,k} \to M \in \mathbb{R}$,

$$\sqrt{k}\left(\hat{\gamma}_k - \gamma\right) \to_d \mathcal{N}\left(\mu, \sigma^2\right)$$

with

$$\mu = \frac{M\tau_M(1-\tilde{\rho})(1-2\tilde{\rho})}{\tilde{\rho}^2 + \tau_M(1-\tilde{\rho})^2(1-2\tilde{\rho})}$$

and

$$\sigma^2 = (1-\gamma)^2\left\{1 + \frac{\tilde{\rho}^2(1-2\tilde{\rho})}{\{\tilde{\rho}^2 + \tau_M(1-\tilde{\rho})^2(1-2\tilde{\rho})\}^2}\right\} + 2\gamma\left\{\xi_0(\gamma)\right.$$
$$\left. + \frac{\tilde{\rho}^2}{(1-\tilde{\rho})^2}\left[\frac{\xi_1(\gamma)}{\tilde{\rho}^2 + \tau_M(1-\tilde{\rho})^2(1-2\tilde{\rho})} + \frac{\xi_1(\gamma)(1-\tilde{\rho})^2(1-2\tilde{\rho})}{\{\tilde{\rho}^2 + \tau_M(1-\tilde{\rho})^2(1-2\tilde{\rho})\}^2}\right]\right\}.$$

Stellenbosch University https://scholar.sun.ac.za

2.3 Simulation study

In this section the characteristics of the ridge regression estimators are illustrated for a variety of different distributions with a simulation study, taking 10 000 repetitions of samples of size $n = 200$. We study the finite sample behavior of $\hat{\gamma}_k$ and also of $\hat{\gamma}_k^+$ in case of Pareto-type distributions, comparing these ridge regression estimators with $GH_{k,n}$, and also with $H_{k,n}$ and $CH_{k,n}$ when $\gamma > 0$. The bias and root means squared error (RMSE) are plotted as a function of k.

The following distributions are used:

- *The Fréchet*(2) *distribution* with $\bar{F}(x) = 1 - \exp\left(-x^{-2}\right)$ for $x > 0$, so that $\gamma = \frac{1}{2}$ and $\tilde{\rho} = -1$.
- *The Burr*$\left(\sqrt{2}, \sqrt{2}\right)$ *distribution* with $\bar{F}(x) = \left(1 + x^{\sqrt{2}}\right)^{-\sqrt{2}}$ for $x > 0$, so that $\gamma = \frac{1}{2}$ and $\tilde{\rho} = -\frac{\sqrt{2}}{2}$.
- *The Log Gamma(2,2) distribution* with $\bar{F}(x) = x^{-2}\ (1 + 2\log x)$ for $x > 1$, so that $\gamma = \frac{1}{2}$ and $\tilde{\rho} = 0$.
- *The Gamma(2,2) distribution* with $\bar{F}(x) = e^{-2x}\ (1 + 2x)$ for $x > 0$, so that $\gamma = 0$ and $\tilde{\rho} = 0$.
- *The Weibull(2) distribution* with $\bar{F}(x) = \exp\left(-x^2\right)$ for $x > 0$, so that $\gamma = 0$ and $\tilde{\rho} = 0$.
- *The Extreme Value Weibull(2) distribution* with $\bar{F}(x) = 1 - \exp\left(-(m - x)^2\right)$ for $x \leq m$, so that $\gamma = -\frac{1}{2}$ in which case $\tilde{\rho} = -\frac{1}{2}$.
- *The Reversed Burr*$\left(\sqrt{2}, \sqrt{2}\right)$ *distribution* with $\bar{F}(x) = \left(1 + (m - x)^{\sqrt{2}}\right)^{-\sqrt{2}}$ for $x \leq m$, so that $\gamma = -\frac{1}{2}$ with $\tilde{\rho} = -\frac{\sqrt{2}}{2}$.

In case of the Pareto-type distributions, see Figures 2.1, 2.2 and 2.3 for the bias and RMSE curves of the ridge regression estimator $\hat{\gamma}_k^+$ are situated between those of the Hill estimator $H_{k,n}$ and the least squares estimator $\hat{\gamma}_k^{\mathrm{LS}+}$. The RMSE curves of $\hat{\gamma}_k^+$ are low and flat over the central k region, where the bias and RMSE of $\hat{\gamma}_k^+$ are close to the least squares plots in general. When applicable, the proposed bias reduced estimator also performs well in comparison with the corrected Hill estimator $CH_{k,n}$ using the Fraga Alves et al. (2003a) estimator of ρ.

Stellenbosch University https://scholar.sun.ac.za

Similar conclusions can be drawn concerning the behavior of $\hat{\gamma}_k$ in comparison with the generalized Hill estimator $GH_{k,n}$ and the corresponding least squares estimator $\hat{\gamma}_k^{\mathrm{LS}}$ (see Figures 2.1 to 2.7). For k up to some value (between 50 and 100 depending on the case) the RMSE of the ridge regression estimator is situated between the smaller $GH_{k,n}$ values and the larger $\hat{\gamma}_k^{\mathrm{LS}}$, while globally the RMSE plots are low and flat staying close to the least squares estimator for larger values of k The bias is globally smallest for the least squares estimators, followed by the ridge regression results. The generalized Hill estimator can lead to substantial bias.

When increasing the sample size the same trends persist while the differences between the least squares and ridge regression estimators become smaller.

Stellenbosch University https://scholar.sun.ac.za

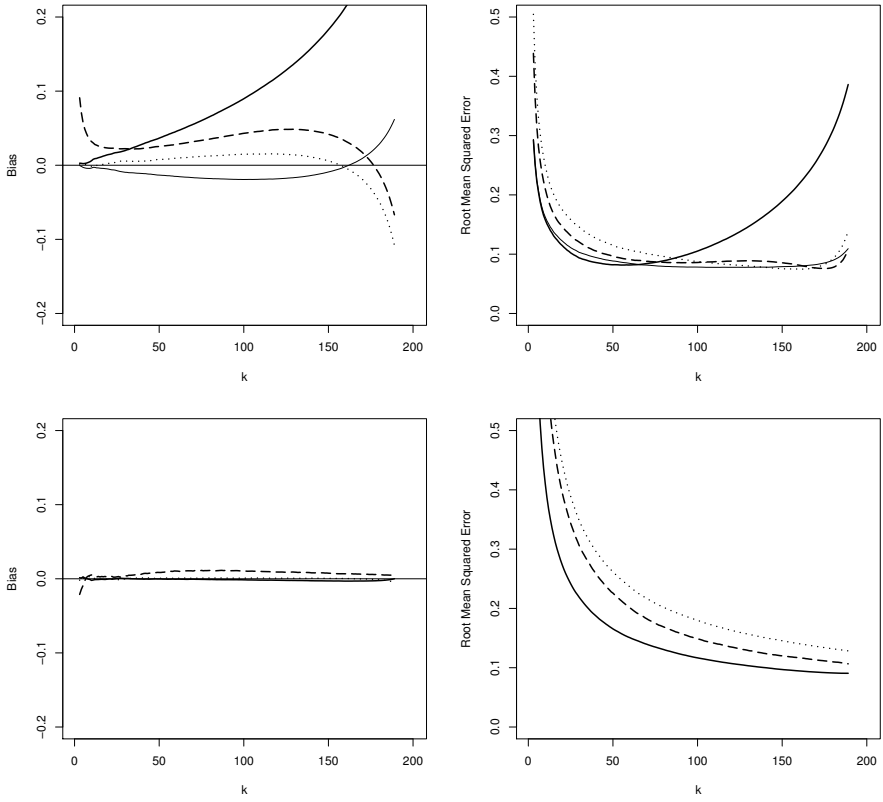

Figure 2.1: Fréchet distribution with $\gamma = \frac{1}{2}$ and $\rho = -1$: bias (left) and RMSE (right) of $H_{k,n}$ (solid), $CH_{k,n}$ (thin solid), $\hat{\gamma}_k^{\mathrm{LS}+}$ (dotted), $\hat{\gamma}_k^{+}$ (dashed) on top; $GH_{k,n}$ (solid), $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted), $\hat{\gamma}_k$ (dashed) on bottom.

Stellenbosch University https://scholar.sun.ac.za

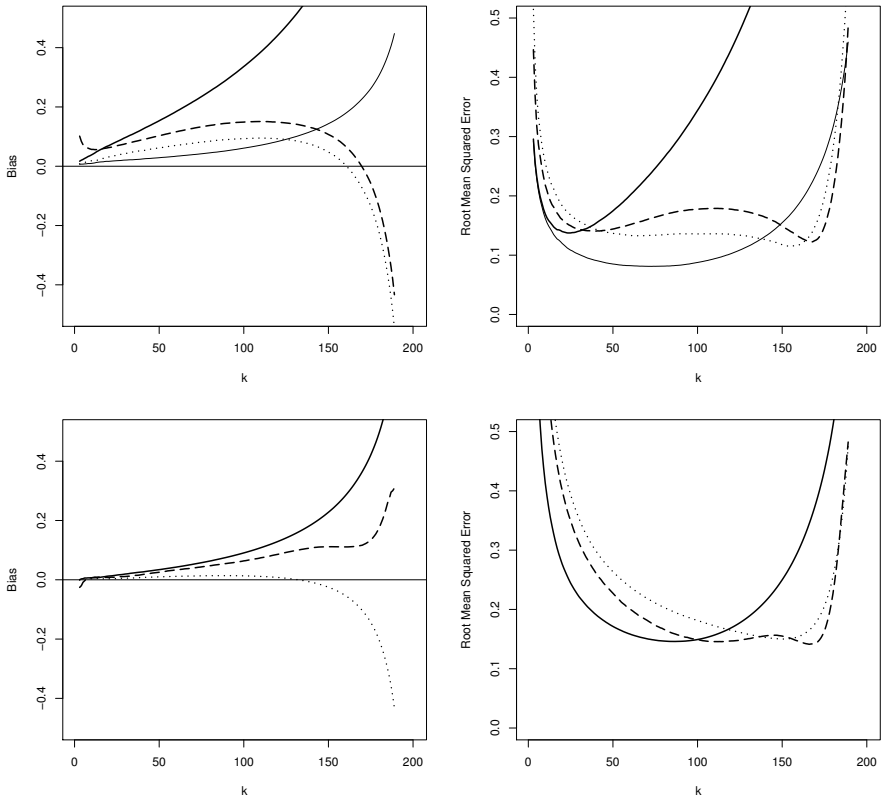

Figure 2.2: Burr distribution with $\gamma = \frac{1}{2}$ and $\rho = -\frac{\sqrt{2}}{2}$: bias (left) and RMSE (right) of $H_{k,n}$ (solid), $CH_{k,n}$ (thin solid), $\hat{\gamma}_k^{\mathrm{LS}+}$(dotted), $\hat{\gamma}_k^{+}$ (dashed) on top; $GH_{k,n}$ (solid), $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted), $\hat{\gamma}_k$ (dashed) on bottom.

Stellenbosch University https://scholar.sun.ac.za

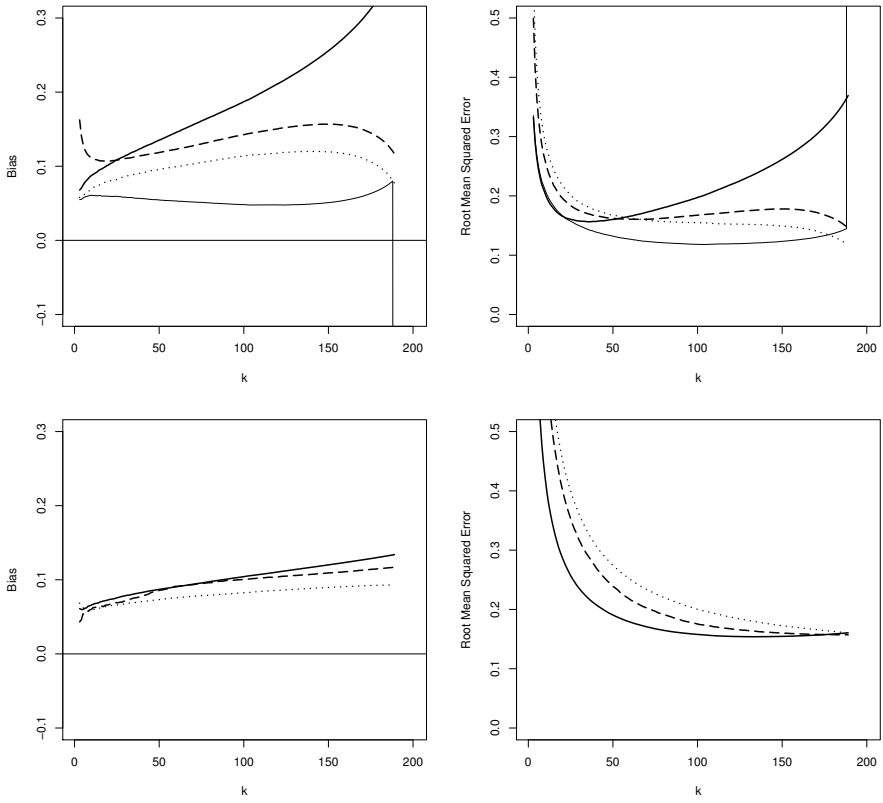

Figure 2.3: Log Gamma distribution with $\gamma = \frac{1}{2}$ and $\rho = 0$: bias (left) and RMSE (right) of $H_{k,n}$ (solid), $CH_{k,n}$ (thin solid), $\hat{\gamma}_k^{\mathrm{LS}+}$(dotted), $\hat{\gamma}_k^{+}$ (dashed) on top; $GH_{k,n}$ (solid), $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted), $\hat{\gamma}_k$ (dashed) on bottom.

Stellenbosch University https://scholar.sun.ac.za

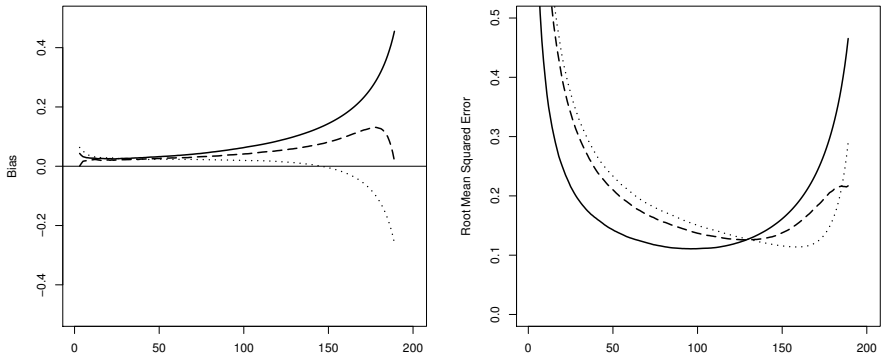

Figure 2.4: Gamma distribution with $\gamma = 0$ and $\tilde{\rho} = -\frac{1}{2}$: bias (left) and RMSE (right) of $GH_{k,n}$ (solid), $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted), $\hat{\gamma}_k$ (dashed).

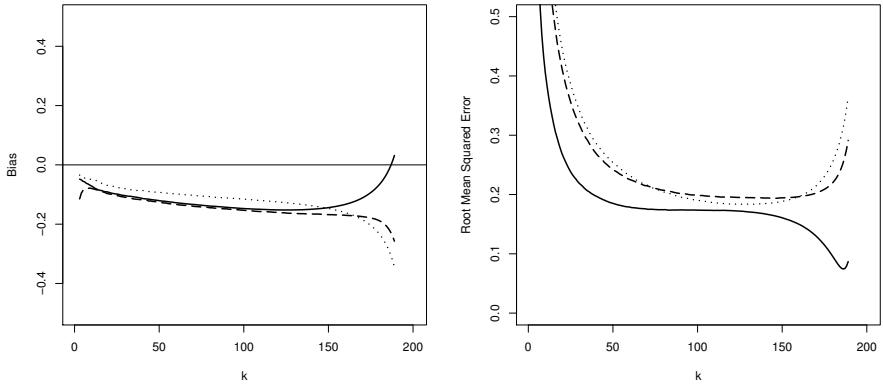

Figure 2.5: Weibull distribution with $\gamma = 0$ and $\tilde{\rho} = 0$: bias (left) and RMSE (right) of $GH_{k,n}$ (solid), $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted), $\hat{\gamma}_k$ (dashed).

Stellenbosch University https://scholar.sun.ac.za

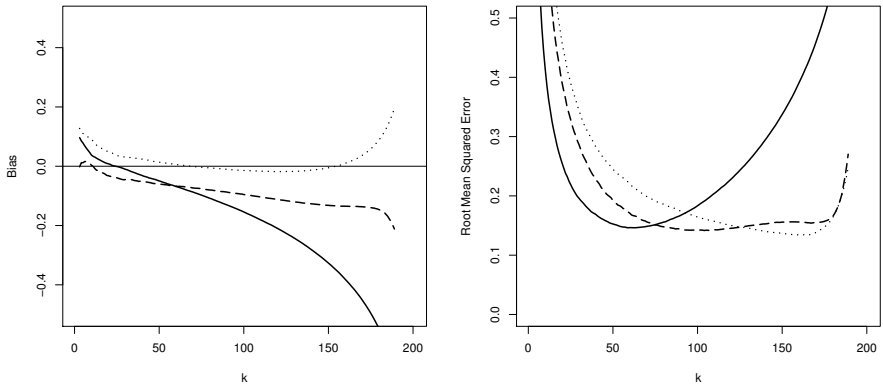

Figure 2.6: Extreme Value Weibull distribution with $\gamma = -\frac{1}{2}$ and $\tilde{\rho} = -\frac{1}{2}$: bias (left) and RMSE (right) of $GH_{k,n}$ (solid), $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted), $\hat{\gamma}_k$ (dashed).

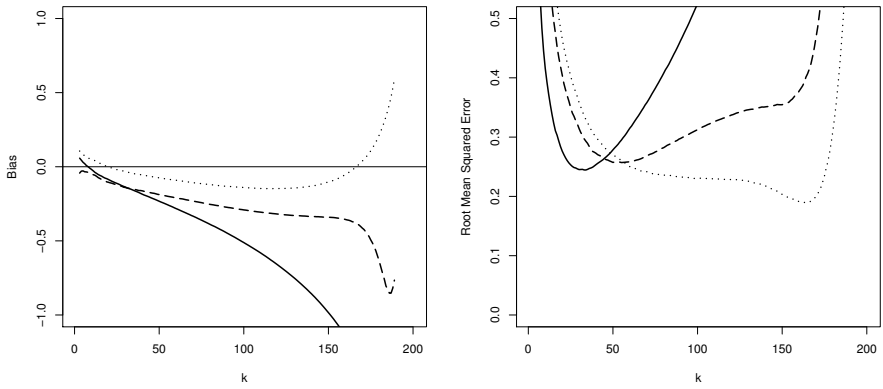

Figure 2.7: Reversed Burr distribution with $\gamma = -\frac{1}{2}$ and $\tilde{\rho} = -\frac{\sqrt{2}}{2}$: bias (left) and RMSE (right) of $GH_{k,n}$ (solid), $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted), $\hat{\gamma}_k$ (dashed).

Stellenbosch University https://scholar.sun.ac.za

2.4 Practical case studies

In this section the ridge regression estimators for the EVI are applied to the daily maximum wind speed data from the US city of Albuquerque as well as to the Secura Belgian reinsurance claim size data. Both data sets were already discussed in Beirlant et al. (2004).

The wind speed data from the city of Albuquerque consists of 6939 observed daily maximum wind speeds measured in miles per hour, which have been rounded off to the nearest integer value. In order to compensate for the loss of accuracy during the data-collection process we have added some small noise to the observations . Given the large size of this data set, we also focus the plots of the EVI estimates to $1 \leq k \leq 2000$ (middle) and to $1 \leq k \leq 500$ (bottom) in Figure 2.8.

The Secura Belgian reinsurance data consists of 371 observed claim sizes in euro amounts which occurred during the years from 1988 to 2001. The EVI estimators of the reinsurance data is illustrated in Figures 2.9 and 2.10. These practical applications further illustrate the advantages of the ridge regression estimators.

The plots of $H_{k,n}$ and $GH_{k,n}$ are hard to interpret, and it is problematic to decide on any specific estimate value for the EVI. The estimators $\hat{\gamma}_k^{\mathrm{LS}+}$ and $\hat{\gamma}_k^{\mathrm{LS}}$ are bias-reduced and therefore more horizontal over a larger region in k, but are more unstable for small values of k. The plots of the ridge regression estimates $\hat{\gamma}_k^{+}$ and $\hat{\gamma}_k$ are stable and flat over most values of k, which simplifies the decision to specify the estimate value for the EVI.

Regarding the daily wind speed data, the stability of the plot of $\hat{\gamma}_k$ is striking. For the largest 2000 observations there is a single distinct level of EVI estimates at -0.1.

In the case of the Secura Belgian reinsurance data a positive EVI becomes apparent in Figure 2.9 and in this case we also applied the Hill estimator $H_{k,n}$, the corrected Hill estimator $CH_{k,n}$ using the Fraga Alves et al. (2003a) estimator of ρ, and the corresponding regression estimator $\hat{\gamma}_k^{+}$ (Figure 10). Here the plot of $\hat{\gamma}_k^{+}$ is very stable, and reveals a single distinct level of EVI estimates at 0.29 over all except the smallest values of k. The plot of $\hat{\gamma}_k$ yields an EVI estimate of around 0.20. In Figure 2.10 we also plot the values of $1/\hat{\tau}_k^{+}$ as

Stellenbosch University https://scholar.sun.ac.za

obtained from (2.12) as a function of k. This plot indicates that for $k > 150$ the results of the Hill estimator are probably too biased since there the Hill and least squares estimator start to diverge. For k between 50 and 150 these two estimators deviate from each other in a limited way.

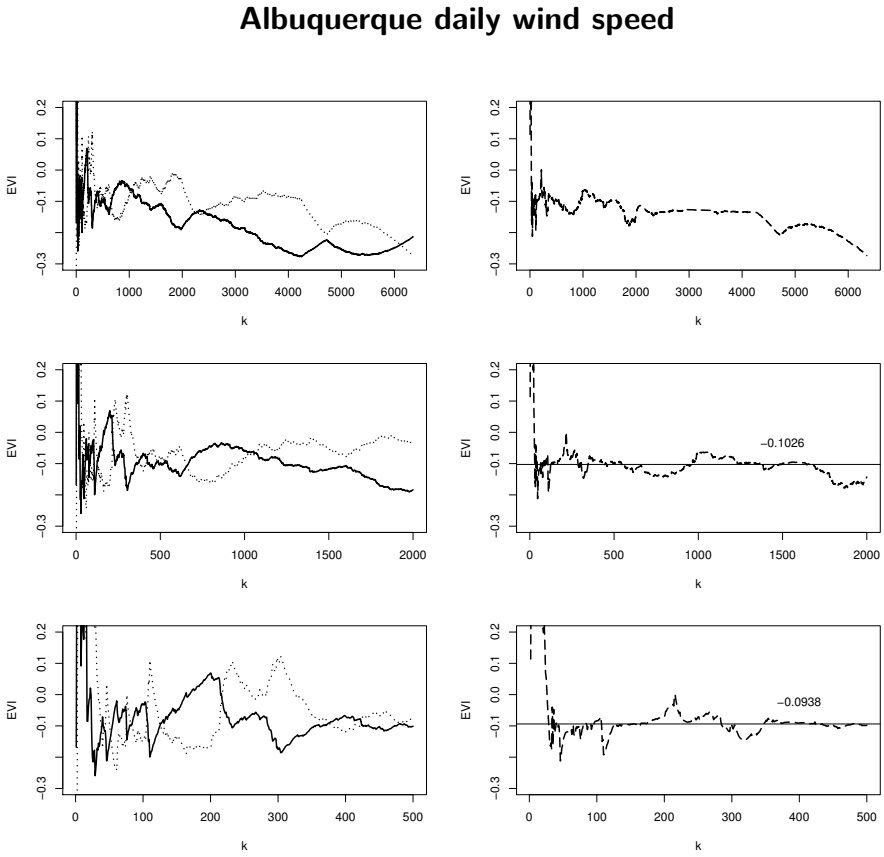

Figure 2.8: Daily wind speed of Albuquerque: $GH_{k,n}$ (solid) and $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted) on the left and $\hat{\gamma}_k$ (dashed) on the right. These estimates are illustrated for the whole dataset (top), focusing on the top 2000 data points (middle) and on the top 500 data points (bottom). Horizontal lines are based on the median of the plotted estimates.

Stellenbosch University https://scholar.sun.ac.za

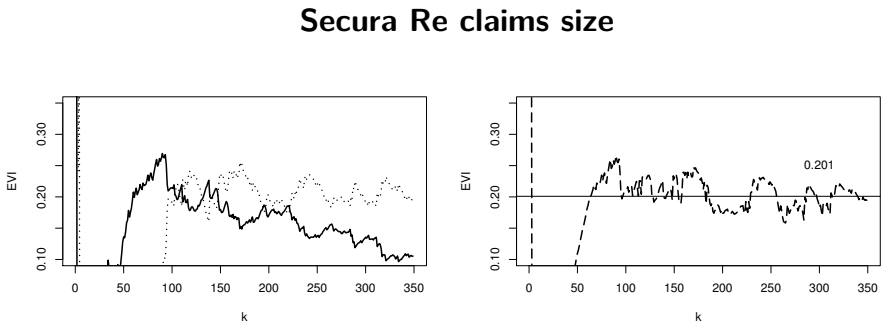

Figure 2.9: Claim sizes of Secura Belgian Reinsurance: $GH_{k,n}$ (thick solid) and $\hat{\gamma}_k^{\mathrm{LS}}$ (dotted) on the left, and $\hat{\gamma}_k$ (dashed) on the right. Horizontal line is based on the median of the plotted estimates.

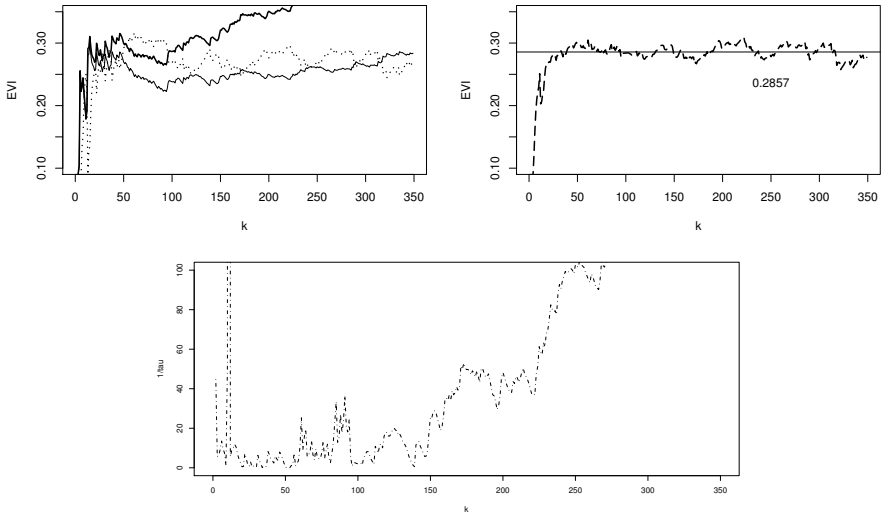

Figure 2.10: Claim sizes of Secura Belgian Reinsurance: $H_{k,n}$ (thick solid), $CH_{k,n}$ (thin solid) and $\hat{\gamma}_k^{\mathrm{LS}+}$ (dotted) on the top left, and $\hat{\gamma}_k^{+}$ (dashed) on the top right. Horizontal line is based on the median of the plotted estimates. Bottom: a plot of $1/\hat{\tau}_k^{+}$ as a function of k.

Stellenbosch University https://scholar.sun.ac.za

2.5 Conclusions

In this paper we studied bias reduced estimators for a real-valued EVI based on a regression model which was proposed in Beirlant et al. (2005). Next to simple least squares estimation we also applied ridge regression. The ridge penalty parameter was chosen using mean squared error computations of the resulting EVI estimator. The ridge regression method was also applied to Pareto-type distributions with positive EVI.

The ridge regression estimators are compromises between the arithmetic mean (which typically has a large bias and a smaller variance) and the least squares estimators (which have a smaller bias but a larger variance). For small values of k, the ridge regression estimators tend to follow the arithmetic mean more closely, while as k becomes larger the ridge regression estimators compare well with the least squares estimator. The characteristics of the ridge estimators are that

- in terms of accuracy, the bias and RMSE of the ridge regression estimators are situated between the arithmetic mean and the least squares estimators,
- their RMSE curves are flat over the central k region,
- the EVI estimate plots are more stable and horizontal than either the $H_{k,n}$, respectively $GH_{k,n}$, and the corresponding simple least squares estimators.

Stellenbosch University https://scholar.sun.ac.za

Chapter 3

Confidence Intervals for Extreme Pareto-type Quantiles

In this paper we revisit the construction of confidence intervals for extreme quantiles of Pareto-type distributions starting from the extreme quantile estimator introduced by Weissman (1978), up to recent bias reduced estimators. A novel asymptotic pivotal quantity is proposed for these quantile estimators which leads to new asymptotic confidence intervals that exhibit more accurate coverage probability. This pivotal quantity also allows for the construction of a saddlepoint approximation, from which a second set of new confidence intervals follows.
The small sample properties and utility of these confidence intervals are studied using simulations and a case study from insurance.

3.1 Introduction

Extreme value methodology is motivated from applications in a variety of fields, including finance, insurance, meteorology and hydrology. The estimation of extreme quantiles as well as the construction of confidence intervals for extreme quantiles then are an important goal in these applications. Although there are several estimators for extreme quantiles available in the literature, very

Stellenbosch University https://scholar.sun.ac.za

little attention has been given to constructing confidence intervals for extreme quantiles. In literature attention is given to asymptotic confidence intervals based on the asymptotic normality of the classically standardized Weissman and bias reduced estimators (see for instance de Haan and Ferreira (2006), Beirlant et al. (2008) and Gomes and Pestana (2007)). In this paper we propose a new pivotal quantity of the quantile estimator which opens the door to the construction of a variety of confidence intervals for extreme Pareto-type quantiles.

Assume that the underlying distribution function F satisfies the max-domain of attraction condition, i.e. assume that the maximum of independent and identically distributed observations $X_1, X_2, \ldots, X_n$ has an asymptotic generalized extreme value distribution:

$$\mathbb{P}\left(a_n^{-1}\left(\max_{i=1,\ldots,n} X_i - b_n\right) \leq y\right) \to G_\gamma(y) = \exp\left(-(1+\gamma y)^{-1/\gamma}\right)$$

for $1+\gamma\, y > 0$ as $n \to \infty$, where $b_n \in \mathbb{R}$, $a_n > 0$ and $\gamma \in \mathbb{R}$ are the location, scale and shape parameters respectively. The extreme value index (EVI) γ is a measure of the tail-heaviness of the distribution of X with a larger value of γ implying a heavier tail of F.

We here confine ourselves to the case $\gamma > 0$. Then a continuous and strictly increasing distribution function F satisfies (2.1) if and only if it belongs to the set of Pareto-type distributions with right tail function given by

$$\bar{F}(x) = 1 - F(x) = P(X > x) = x^{-1/\gamma}\, \ell(x),$$

where ℓ is a slowly varying function at infinity, i.e.

$$\frac{\ell(tx)}{\ell(t)} \to 1 \text{ as } t \to \infty \text{ for every } x > 1.$$

This means that the exceedances $Y_t = \frac{X}{t}$ given $X > t$ over a high threshold t, are approximately Pareto distributed with shape parameter $\frac{1}{\gamma}$. We further consider the tail quantile function T defined as

$$T(x) = Q\left(1 - \frac{1}{x}\right) \quad \text{for } x > 1,$$

where Q denotes the quantile function defined as the inverse function of F. The max-domain of attraction condition (2.2) is then equivalent to

$$T(x) = x^\gamma \ell_T(x),$$

Stellenbosch University https://scholar.sun.ac.za

where ℓ_T is a slowly varying function at infinity, i.e. ℓ_T also satisfies the condition (2.3) and is the de Bruijn conjugate of ℓ.

We next review the most commonly used extreme quantile estimators for Pareto-type distributions. To this end we introduce the weighted log-spacings Z_j as

$$Z_j = j \log \frac{X_{n-j+1,n}}{X_{n-j,n}} \quad \text{for } j = 1, \ldots, k$$

where $X_{1,n} \leq X_{2,n} \leq \ldots \leq X_{n,n}$ denote the ordered data.

Weissman-type quantile estimators. From the Rényi representation of exponential order statistics it follows that if ℓ in (2.2) is constant

$$Z_j =_d \gamma E_j, \tag{3.1}$$

where $E_1, E_2, \ldots, E_n$ are independent standard exponentially distributed random variables. As detailed in Theorem 4.1 in Beirlant et al. (2004) it follows that this representation still holds approximately for $j = 1, \ldots, k$ with $k, n \to \infty$ and $k/n \to 0$. From this, the famous Hill (1975) estimator for γ follows readily:

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} Z_j.$$

Weissman (1978) introduced a basic estimation method for an extreme quantile $Q(1-p)$ with p small of a Pareto-type tail:

$$\hat{Q}_k(1-p; \hat{\gamma}_k) = X_{n-k,n} \left(\frac{k}{np} \right)^{\hat{\gamma}_k} \tag{3.2}$$

where $\hat{\gamma}_k$ is any suitable estimator for γ. This estimator follows from (2.7) as then, for $1/p, n/k \to \infty$,

$$\frac{T(1/p)}{T(n/k)} \sim (\frac{1/p}{n/k})^{\gamma} = (\frac{k}{np})^{\gamma},$$

and estimating $T(n/k) = Q(1 - \frac{k}{n})$ by $X_{n-k,n}$. Originally, Weissman (1978) proposes to use the Hill estimator $H_{k,n}$ as $\hat{\gamma}_k$, in which case we use the notation $\hat{Q}_k^{\mathrm{H}}(1-p) = \hat{Q}_k(1-p; H_{k,n})$.

Stellenbosch University https://scholar.sun.ac.za

The accuracy of $\hat{Q}_k(1-p;\hat{\gamma}_k)$ is known to depend in first order on the accuracy of the EVI estimator $\hat{\gamma}_k$. For instance, in case of $\hat{Q}_k^{\mathrm{H}}(1-p)$, as $H_{k,n}$ increases in bias and decreases in variance as the number of top observations k increases, so does the quantile estimator. The level of bias of $H_{k,n}$ depends on the extent to which the distribution of the spacings Z_j differs from the exponential distribution, and therefore depends on the rate of convergence in (2.3) when applied to ℓ_T.

Bias-reduced quantile estimators. In order to assess the bias of estimators of γ and of extreme quantiles $Q(1-p)$, we need more than just the first order condition as (2.3) for ℓ_T. A convenient refinement is that there exists some positive regularly varying function b with index $\rho<0$, i.e. $b(x)=x^{\rho}\ell_b(x)$ for some slowly varying ℓ_b, such that for all $x>0$

$$\frac{\log \ell_T(tx)-\log \ell_T(t)}{b(t)} \to h_\rho(x) := \frac{x^\rho-1}{\rho} \qquad \text{as } t\to\infty. \tag{3.3}$$

When discussing bias reduction it is useful to consider the Hall class of heavy tailed distributions (see Hall (1982)) satisfying (3.3), with $\rho<0$, $C>0$ and $D\neq 0$,

$$T(x)=Cx^{\gamma}\left(1+Dx^{\rho}(1+o(1))\right), \text{ as } x\to\infty. \tag{3.4}$$

Here for mathematical convenience we use the reparametrization $D=\frac{\gamma}{\rho}\beta$ with $\beta\neq 0$, so that $b(x)=\gamma\beta x^{\rho}(1+o(1))$.

Under (3.4) Feuerverger and Hall (1999) and Beirlant et al. (1999) proposed an exponential regression model for Z_j $(j=1,2,\ldots,k)$ where the EVI γ is the intercept, $b_{n,k}=b\left(\frac{n}{k}\right)$ is the gradient and $c_{j,k}=\left(\frac{j}{k+1}\right)^{-\rho}$ is the covariate (see Theorem 4.1 in Beirlant et al. (2004)):

$$Z_j =_d (\gamma + b_{n,k}\, c_{j,k})\; E_j + o_p(b_{n,k}) \quad \text{for } j=1,\ldots,k. \tag{3.5}$$

Under (3.4) and based on the representation (3.5), Matthys et al. (2004) introduced the following bias-reduced estimator for the quantile function:

$$\hat{Q}_k(1-p;\hat{\gamma}_k,\hat{b}_{n,k}) = X_{n-k,n}\,\exp\left\{\hat{\gamma}_k\, h_0\left(\frac{k}{np}\right)+\hat{b}_{n,k}\, h_{\hat{\rho}}\left(\frac{k}{np}\right)\right\}, \tag{3.6}$$

assuming that $\hat{\gamma}_k\, h_0\left(\frac{k}{np}\right)+\hat{b}_{n,k}\, h_{\hat{\rho}}\left(\frac{k}{np}\right)>0$ so that $\hat{Q}_k(1-p;\hat{\gamma}_k,\hat{b}_{n,k})>X_{n-k,n}$, with $h_0(x)=\log x$. Here $\hat{b}_{n,k}$ denotes some suitable estimator of $b_{n,k}$, for instance based on solving the regression model (3.5). Under (3.4) estimators $\hat{\rho}=\hat{\rho}_{k^*}$ of the second order parameter ρ are available based on a number k^*

Stellenbosch University https://scholar.sun.ac.za

of top data with $\sqrt{k^*}b(n/k^*) \to \infty$. For instance, Fraga Alves et al. (2003b) proposed

$$\hat{\rho} = -\frac{3(T-1)}{T-3} \tag{3.7}$$

where $T = \frac{M_1 - M_2}{M_2 - M_3}$, $M_m = \left(\frac{\frac{1}{k^*} \sum_{j=1}^{k^*} (\log X_{n-j+1,n} - \log X_{n-m,n})^m}{m!} \right)^{\frac{1}{m}}$ for $m = 1, 2, 3$ and $k^* = \lfloor n^{.99} \rfloor$. See Remark 3.2 in Gomes et al. (2007) stating that $\hat{\rho} - \rho = o_p(1/(\log n))$ under (3.4). Below we denote the estimated value of $c_{j,k}$ with $d_{j,k}$

$$d_{j,k} := (\frac{j}{k+1})^{-\hat{\rho}}.$$

Buitendag et al. (2019) proposed ridge regression type estimators for γ and $b_{n,k}$ based on the representation (3.5) given by

$$\hat{\gamma}_k(\tau) = H_{k,n} - \bar{d}_k \, \hat{b}_{n,k}(\tau) \tag{3.8}$$

$$\hat{b}_{n,k}(\tau) = \frac{\sum_{j=1}^{k} (d_{j,k} - \bar{d}_k) \, Z_j}{\sum_{j=1}^{k} (d_{j,k} - \bar{d}_k)^2 + k\,\tau} \tag{3.9}$$

where $\bar{d}_k = \frac{1}{k} \sum_{j=1}^{k} d_{j,k}$ and $\tau > 0$ is the ridge regression penalty parameter. When $\tau \to \infty$, $\hat{\gamma}_k(\tau)$ reduces to the Hill estimator $H_{k,n}$. Choosing $\tau = 0$, (3.9) and (3.8) lead to the least squares estimators

$$\hat{\gamma}_k^{\mathrm{LS}} = \hat{\gamma}_k(0) \qquad \text{and} \qquad \hat{b}_{n,k}^{\mathrm{LS}} = \hat{b}_{n,k}(0), \tag{3.10}$$

proposed in Feuerverger and Hall (1999) and Beirlant et al. (1999).

In Buitendag et al. (2019) the penalty parameter τ is chosen as $\tau_k = \frac{\gamma^2}{k\, b_{n,k}^2} = \frac{1}{k\beta^2 (\frac{n}{k})^{2\rho}}$ which minimizes the asymptotic MSE of the EVI estimator $\hat{\gamma}_k(\tau)$. In order to estimate τ_k, under (3.4), the following estimator for $\hat{\tau}_k$ was proposed:

$$\hat{\tau}_k = \frac{1}{k\,\hat{\beta}^2 \, \left(\frac{n}{k}\right)^{2\hat{\rho}}}, \tag{3.11}$$

with

$$\hat{\beta} = \left(\frac{k^\star}{n}\right)^{\hat{\rho}} \frac{\left(\sum_{j=1}^{k^\star} d_{j,k^\star}\right)\left(\sum_{j=1}^{k^\star} Z_j\right) - k^\star \sum_{j=1}^{k^\star} d_{j,k^\star} Z_j}{\left(\sum_{j=1}^{k^\star} d_{j,k^\star}\right)\left(\sum_{j=1}^{k^\star} d_{j,k^\star} Z_j\right) - k^\star \sum_{j=1}^{k^\star} d_{j,k^\star}^2 Z_j}, \tag{3.12}$$

Stellenbosch University https://scholar.sun.ac.za

a consistent estimator of β with $\sqrt{k}(\hat{\beta}-\beta)=O_p(1)$ under (3.4) (see Gomes and Martins (2002)).

The ridge regression estimators for γ and $b_{n,k}$ follow from (3.8) and (3.9) by choosing τ as (3.11):

$$\hat{\gamma}_k^{\text{ridge}}=\hat{\gamma}_k\left(\hat{\tau}_k\right) \qquad \text{and} \qquad \hat{b}_{n,k}^{\text{ridge}}=\hat{b}_{n,k}\left(\hat{\tau}_k\right). \tag{3.13}$$

Caeiro et al. (2005) proposed a corrected Hill estimator to estimate γ and $b_{n,k}$ using the above statistics $\hat{\beta}$ and $\hat{\rho}$:

$$\begin{aligned}\hat{\gamma}_k^{\text{CH}}&=\left(1-\frac{\hat{\beta}\,\left(\frac{n}{k}\right)^{\hat{\rho}}}{1-\hat{\rho}}\right)H_{k,n}\\ \hat{b}_{n,k}^{\text{CH}}&=\hat{\gamma}_k^{CH}\,\hat{\beta}\,\left(\frac{n}{k}\right)^{\hat{\rho}}.\end{aligned} \tag{3.14}$$

From (3.6), using the above mentioned different estimators of $(\gamma, b_{n,k})$ we obtain three bias-reduced estimators:

- ridge regression $\qquad \hat{Q}_k^{\text{ridge}}(1-p)=\hat{Q}_k(1-p;\hat{\gamma}_k^{\text{ridge}},\hat{b}_{n,k}^{\text{ridge}}),$
- least squares regression $\qquad \hat{Q}_k^{\text{LS}}(1-p)=\hat{Q}_k(1-p;\hat{\gamma}_k^{\text{LS}},\hat{b}_{n,k}^{\text{LS}}),$
- corrected Hill $\qquad \hat{Q}_k^{\text{CH}}(1-p)=\hat{Q}_k(1-p;\hat{\gamma}_k^{\text{CH}},\hat{b}_{n,k}^{\text{CH}}).$

Note that $\hat{Q}_k^{\text{CH}}(1-p)$ was already discussed in Beirlant et al. (2008) and that

$$\hat{Q}_k^{\text{H}}(1-p)=\hat{Q}(1-p;H_{k,n},0).$$

In the next section we review the asymptotic confidence intervals based on the classical pivot

$$\sqrt{k}\,\frac{\log\hat{Q}_k^{\text{H}}(1-p)-\log Q(1-p)}{\gamma\,h_0(\frac{k}{np})}$$

which is known to be asymptotically normally distributed (see for instance de Haan and Ferreira (2006)). We then introduce a new pivotal quantity whose asymptotic normal distribution does not depend on the EVI γ. In section 3.3 we explain how the distribution of this new pivot can be approximated using saddlepoint methods leading to a second set of confidence intervals. In a final section we compare the different confidence intervals using simulations and a case study from non-life insurance.

Stellenbosch University https://scholar.sun.ac.za

3.2 A new pivotal quantity and asymptotic normal confidence intervals

Asymptotic results for the original quantile estimator $\hat{Q}_k^{\mathrm{H}}(1-p)$ go back to de Haan and Rootzén (1993) and are reviewed for instance in de Haan and Ferreira (2006): assuming $k, n \to \infty$ with $k/n \to 0$ such that $\sqrt{k}\, b_{n,k} \to M$ and $p = p_n$ satisfies $k/(np_n) \to \infty$, then under (3.4)

$$\sqrt{k}\, \frac{\log \hat{Q}_k^{\mathrm{H}}(1-p) - \log Q(1-p)}{\gamma\, h_0(\frac{k}{np})} \to_d \mathcal{N}\left(\frac{M}{\gamma(1-\rho)}, 1\right).$$

More generally, from Beirlant et al. (2008) and Buitendag et al. (2019) it follows also that

$$T_{k,n}^{\bullet} := \sqrt{k}\, \frac{\log \hat{Q}_k^{\bullet}(1-p) - \log Q(1-p)}{\gamma\, h_0(\frac{k}{np})} = \sqrt{k}\left(\frac{\hat{\gamma}_k^{\bullet}}{\gamma} - 1\right) + o_p(1) \to_d \mathcal{N}\left(\mu, \sigma^2\right) \tag{3.15}$$

where $\hat{Q}_k^{\bullet}$ refers to one of the four quantile estimators $\hat{Q}_k^{H}$, $\hat{Q}_k^{\mathrm{LS}}$, $\hat{Q}_k^{\mathrm{ridge}}$ or $\hat{Q}_k^{\mathrm{CH}}$ based on the corresponding estimators $(\hat{\gamma}_k^{\bullet}, \hat{b}_{n,k}^{\bullet})$ of $(\gamma, b_{n,k})$. The limiting parameters μ and σ^2 are given in Table 3.1, with $\tau = \frac{\gamma^2}{M^2} = \lim \tau_k$ as $k, n \to \infty$ where the limit is taken over $k/n \to 0$ and $\sqrt{k} b_{n,k} \to M$.

Moreover in case of the bias reduced estimators $\hat{Q}_k^{\mathrm{LS}}$, $\hat{Q}_k^{\mathrm{ridge}}$, $\hat{Q}_k^{\mathrm{CH}}$ these limiting parameters (μ, σ^2) only depend on the second order parameters ρ, β, M and τ but not on γ directly, and hence $T_{k,n}^{\bullet}$ can be considered as an asymptotic pivot.

The classical standardization (3.15) leads to the following two-sided $(1-\alpha)100\%$ asymptotic confidence intervals for $Q(1-p)$

$$\left[\hat{Q}_k^{\bullet}(1-p)\, e^{-\hat{\gamma}_k^{\bullet} h_0(\frac{k}{np})\left(\frac{\hat{\mu}^{\bullet} + \hat{\sigma}^{\bullet} z_{1-\alpha/2}}{\sqrt{k}}\right)},\ \hat{Q}_k^{\bullet}(1-p)\, e^{-\hat{\gamma}_k^{\bullet} h_0(\frac{k}{np})\left(\frac{\hat{\mu}^{\bullet} + \hat{\sigma}^{\bullet} z_{\alpha/2}}{\sqrt{k}}\right)}\right], \tag{3.16}$$

where $z_{\alpha/2}$ and $z_{1-\alpha/2}$ are the $\alpha/2$ and $1-\alpha/2$ quantiles of the standard normal distribution, with $\hat{\mu}^{\bullet}$ and $\hat{\sigma}^{\bullet}$ following from Table 3.1 replacing ρ, τ by the estimators defined in (3.7) and (3.11), and M by $\sqrt{k}\hat{b}_{n,k}^{\bullet}$ as defined in (3.10), (3.13) and (3.14). Note that consistency of $\hat{\mu}^{\bullet}/\sqrt{k}$ and $\hat{\sigma}^{\bullet}/\sqrt{k}$ follows from the consistency of $\hat{b}_{n,k}^{\bullet}$, $\hat{\rho}$ and $\hat{\gamma}_k^{\bullet}$.

Stellenbosch University https://scholar.sun.ac.za

Here we restrict attention to the substitution of γ with $\hat{\gamma}_k^\bullet$ in the denominator of $T_{k,n}^\bullet$, i.e. with the same estimator of γ as in the quantile estimator.

Table 3.1: Asymptotic mean and variance of $T_{k,n}^\bullet$ and $S_{k,n}^\bullet$ corresponding to each quantile estimator

Quantile estimator ($\hat{Q}_k^\bullet$)	Asymptotic mean (μ)	Asymptotic variance (σ^2)
$\hat{Q}_k^H$	$\frac{M}{\gamma(1-\rho)}$	1
$\hat{Q}_k^{\mathrm{LS}}$	0	$(\frac{1-\rho}{\rho})^2$
$\hat{Q}_k^{\mathrm{ridge}}$	$\frac{M\tau(1-\rho)(1-2\rho)}{\rho^2+\tau(1-\rho)^2(1-2\rho)}$	$1+\frac{\rho^2(1-2\rho)}{\{\rho^2+\tau(1-\rho)^2(1-2\rho)\}^2}$
$\hat{Q}_k^{\mathrm{CH}}$	0	1

The classically standardized statistics $T_{k,n}^\bullet$ in fact are studentized type statistics since the denominator contains the EVI which leads to the imputation of an estimator $\hat{\gamma}_k$ when constructing confidence intervals. To avoid this we propose a new asymptotic pivotal quantity

$$S_{k,n}^\bullet := \sqrt{k}\,\frac{\log \hat{Q}_k^\bullet(1-p) - \log Q(1-p)}{\log Q(1-p) - \log X_{n-k,n}}. \tag{3.17}$$

We first derive asymptotic representations of the statistics of the type (3.17). We use the notation

$$\kappa_\rho\left(\frac{k}{np}\right) = \frac{h_\rho\left(\frac{k}{np}\right)}{h_0\left(\frac{k}{np}\right)} = \frac{\left(\frac{k}{np}\right)^\rho - 1}{\rho\,\log\left(\frac{k}{np}\right)}.$$

Note that $\kappa_\rho\left(\frac{k}{np}\right) \sim -\frac{1}{\rho h_0(k/(np))} \to 0$ as $k/(np) \to \infty$.

Theorem 3.1. *Assume $k, n \to \infty$ with $k/n \to 0$ such that $b_{n,k} = O(k^{-1/2})$ and that $p = p_n$ satisfies $k/(np_n) \to \infty$. Then under* (3.4)

$$\frac{S_{k,n}^\bullet}{\sqrt{k}} = \frac{\hat{\gamma}_k^\bullet + \hat{b}_{n,k}^\bullet\,\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma + b_{n,k}\,\kappa_\rho\left(\frac{k}{np}\right)}\left(1 - \frac{\log(\frac{n}{k}\,U_{k+1,n})}{h_0(\frac{k}{np})}\right) - 1 + o_p\left(\frac{1}{\sqrt{k}\,h_0(\frac{k}{np})}\right). \tag{3.18}$$

Stellenbosch University https://scholar.sun.ac.za

Proof We first consider the denominator $\log Q(1-p) - \log X_{n-k,n}$ of $k^{-1/2}S^{\bullet}_{k,n}$. Since $X_{n-k,n} \stackrel{d}{=} T(U^{-1}_{k+1,n})$ where $\{U_{j,n}, j = 1, \ldots, n\}$ denote the order statistics from a random uniform (0,1) sample of size n, it follows that as $k, n \to \infty$ with $\frac{k}{n} \to 0$ and $\frac{k}{np} \to \infty$

$$
\begin{aligned}
&\log Q(1-p) - \log X_{n-k,n} \\
&= \log T\left(\frac{1}{p}\right) - \log T\left(\frac{1}{U_{k+1,n}}\right) \\
&= -\gamma \log p + \log \ell_T\left(\frac{1}{p}\right) + \gamma \log U_{k+1,n} - \log \ell_T\left(\frac{1}{U_{k+1,n}}\right) \\
&= \gamma\left(h_0(\frac{k}{np}) + \log U_{k+1,n} - \log \frac{k}{n}\right) + \log \frac{\ell_T\left(\frac{U_{k+1,n}}{p}\,\frac{1}{U_{k+1,n}}\right)}{\ell_T\left(\frac{1}{U_{k+1,n}}\right)} \\
&= \gamma\left(h_0(\frac{k}{np}) + \log(\frac{n}{k}\,U_{k+1,n})\right) + b\left(\frac{1}{U_{k+1,n}}\right)\, h_\rho\left(\frac{U_{k+1,n}}{p}\right)(1 + o_p(1)) \\
&= \gamma h_0\left(\frac{k}{np}\right) + \gamma\, \log(\frac{n}{k}\,U_{k+1,n}) + b_{n,k}\, h_\rho\left(\frac{k}{np}\right)(1 + o_p(1)), \qquad (3.19)
\end{aligned}
$$

since it follows from Smirnov's lemma (see for instance Lemma 2.2.3 in de Haan and Ferreira (2006)) that $\sqrt{k}\left(\frac{n}{k}U_{k+1,n} - 1\right) \to_d \mathcal{N}(0,1)$ so that $b\left(\frac{1}{U_{k+1,n}}\right) = b_{n,k}\,(1 + o_p(1))$ and $h_\rho\left(\frac{U_{k+1,n}}{p}\right) \sim h_\rho\left(\frac{k}{np}\right) \to -\frac{1}{\rho}$ as $k, n \to \infty$ with $\frac{k}{n} \to 0$ and $\frac{k}{np} \to \infty$.

Stellenbosch University https://scholar.sun.ac.za

Finally combining (3.6) and (3.19) we obtain

$$\begin{aligned}
&\frac{\log \hat{Q}_k(1-p;\hat{\gamma}_k^\bullet,\hat{b}_{n,k}^\bullet)-\log Q(1-p)}{\log Q(1-p)-\log X_{n-k,n}}\\
&=\frac{\log \hat{Q}_k(1-p;\hat{\gamma}_k^\bullet,\hat{b}_{n,k}^\bullet)-\log X_{n-k,n}}{\log Q(1-p)-\log X_{n-k,n}}-1\\
&=\frac{\hat{\gamma}_k^\bullet+\hat{b}_{n,k}^\bullet\,\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\left(\gamma+b_{n,k}\,\kappa_\rho\left(\frac{k}{np}\right)\right)\left(1+\frac{\gamma}{\gamma+b_{n,k}\,\kappa_\rho\left(\frac{k}{np}\right)}\frac{\log(\frac{n}{k}\,U_{k+1,n})+o_p(b_{n,k})}{h_0(\frac{k}{np})}\right)}-1\\
&=\frac{\hat{\gamma}_k^\bullet+\hat{b}_{n,k}^\bullet\,\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma+b_{n,k}\,\kappa_\rho\left(\frac{k}{np}\right)}\left(1-\frac{\log(\frac{n}{k}\,U_{k+1,n})+o_p(b_{n,k})}{h_0(\frac{k}{np})}\,(1+o_p(1))\right)-1\\
&=\frac{\hat{\gamma}_k^\bullet+\hat{b}_{n,k}^\bullet\,\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma+b_{n,k}\,\kappa_\rho\left(\frac{k}{np}\right)}\left(1-\frac{\log(\frac{n}{k}\,U_{k+1,n})}{h_0(\frac{k}{np})}\right)-1+\frac{o_p\left(k^{-\frac{1}{2}}\right)+o_p(b_{n,k})}{h_0(\frac{k}{np})}\\
&=\frac{\hat{\gamma}_k^\bullet+\hat{b}_{n,k}^\bullet\,\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma+b_{n,k}\,\kappa_\rho\left(\frac{k}{np}\right)}\left(1-\frac{\log(\frac{n}{k}\,U_{k+1,n})}{h_0(\frac{k}{np})}\right)-1+o_p\left(\frac{1}{\sqrt{k}\,h_0(\frac{k}{np})}\right).
\end{aligned}$$

since $b_{n,k}=O\left(k^{-\frac{1}{2}}\right)$.

From the preceding Theorem the asymptotic distribution of $S_{k,n}^\bullet$ follows, leading to new confidence intervals for $Q(1-p)$. It turns out that $S_{k,n}^\bullet$ has the same limiting distribution as $T_{k,n}^\bullet$.

Corollary 3.1. *Assume $k,n\to\infty$ with $k/n\to 0$ such that $\sqrt{k}b_{n,k}\to M$ finite, and that $p=p_n$ satisfies $k/(np_n)\to\infty$. Then we have under (3.4) that*

$$S_{k,n}^\bullet\to_d\mathcal{N}\left(\mu,\sigma^2\right) \tag{3.20}$$

where μ and σ^2 are given in Table 3.1.

Stellenbosch University https://scholar.sun.ac.za

Proof Under (3.4) it follows from Theorem 1, (3.15) and the fact that $\sqrt{k}(\hat{b}^{\bullet}_{n,k} - b_{n,k}) = O_p(1)$ in each of the three considered estimators (see Addendum B in the Appendix),

$$\begin{aligned}
&\sqrt{k}\frac{\log \hat{Q}_k(1-p;\hat{\gamma}^{\bullet}_k,\hat{b}^{\bullet}_{n,k}) - \log Q(1-p)}{\log Q(1-p) - \log X_{n-k,n}} \\
=&\sqrt{k}\left(\frac{\hat{\gamma}^{\bullet}_k + \hat{b}^{\bullet}_{n,k}\,\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma + b_{n,k}\,\kappa_{\rho}\left(\frac{k}{np}\right)}\left(1 - \frac{\log(\frac{n}{k}\,U_{k+1,n})}{h_0(\frac{k}{np})}\right) + \frac{o_p\left(k^{-\frac{1}{2}}\right)}{h_0(\frac{k}{np})} - 1\right) \\
=&\frac{\sqrt{k}\,(\hat{\gamma}^{\bullet}_k - \gamma) + O_p\left(\left(h_0(\frac{k}{np})\right)^{-1}\right)}{\gamma\,(1+o_p(1))} + o_p\left(\left(h_0(\frac{k}{np})\right)^{-1}\right) \\
=&\sqrt{k}\,\left(\frac{\hat{\gamma}^{\bullet}_k}{\gamma} - 1\right) + O_p\left(\left(h_0(\frac{k}{np})\right)^{-1}\right) \\
\to_d&\mathcal{N}\left(\mu,\sigma^2\right),
\end{aligned}$$

where we used the consistency of $\hat{\rho}$ in step 3 in the above derivation.

From (3.20) we obtain the following two-sided $(1-\alpha)100\%$ confidence interval for $Q(1-p)$ assuming $\hat{Q}^{\bullet}_k(1-p) > X_{n-k,n}$:

$$\left[\hat{Q}^{\bullet}_k(1-p)^{\frac{1}{1+y^{\bullet}_{1-\alpha/2}}}\,X_{n-k,n}^{\frac{(\hat{\mu}^{\bullet}+\hat{\sigma}^{\bullet}z_{1-\alpha/2})/\sqrt{k}}{1+y^{\bullet}_{1-\alpha/2}}},\ \hat{Q}^{\bullet}_k(1-p)^{\frac{1}{1+y^{\bullet}_{\alpha/2}}}\,X_{n-k,n}^{\frac{(\hat{\mu}^{\bullet}+\hat{\sigma}^{\bullet}z_{\alpha/2})/\sqrt{k}}{1+y^{\bullet}_{\alpha/2}}}\right] \tag{3.21}$$

where $y^{\bullet}_{\alpha} = (\hat{\mu}^{\bullet} + \hat{\sigma}^{\bullet}z_{\alpha})/\sqrt{k}$.

Remark 3.1. Let $U^{\mathrm{S}}_k(p)$ and $L^{\mathrm{S}}_k(p)$ denote the upper and lower confidence bounds in (3.21), and similarly let $U^{\mathrm{T}}_k(p)$ and $L^{\mathrm{T}}_k(p)$ denote the upper and lower confidence bounds in (3.16). It follows with (3.6) that

$$\frac{U^{\mathrm{T}}_k(p)}{L^{\mathrm{T}}_k(p)} = \exp\left(2\hat{\gamma}^{\bullet}_k\hat{\sigma}^{\bullet}z_{1-\alpha/2}\frac{h_0\left(\frac{k}{np}\right)}{\sqrt{k}}\right),$$

Stellenbosch University https://scholar.sun.ac.za

and

$$\frac{U_k^{\mathrm{S}}(p)}{L_k^{\mathrm{S}}(p)} = \exp\left(\frac{2}{\sqrt{k}}\left\{\hat{\gamma}_k^{\bullet} h_0\left(\frac{k}{np}\right) + \hat{b}_{k,n}^{\bullet} h_{\hat{\rho}}\left(\frac{k}{np}\right)\right\} \frac{z_{1-\alpha/2}\hat{\sigma}^{\bullet}}{(1+\frac{\hat{\mu}^{\bullet}}{\sqrt{k}})^2 - \frac{(\hat{\sigma}^{\bullet})^2 z_{1-\alpha/2}^2}{k}}\right)$$

$$= \exp\left(\frac{2z_{1-\alpha/2}\hat{\sigma}^{\bullet}}{\sqrt{k}}\left(\hat{\gamma}_k^{\bullet} h_0\left(\frac{k}{np}\right) + \hat{b}_{k,n}^{\bullet} h_{\hat{\rho}}\left(\frac{k}{np}\right)\right)\right.$$

$$\left.\left(1 - 2\frac{\hat{\mu}^{\bullet}}{\sqrt{k}} + \frac{3(\hat{\mu}^{\bullet})^2 + (\hat{\sigma}^{\bullet})^2 z_{1-\alpha/2}^2}{k}(1+o(1))\right)\right).$$

Denote the ratio of the relative lengths of the two confidence intervals as

$$R_k(p) = \frac{U_k^{\mathrm{S}}(p) \,/\, L_k^{\mathrm{S}}(p)}{U_k^{\mathrm{T}}(p) \,/\, L_k^{\mathrm{T}}(p)}.$$

Then as $k, n \to \infty$, $k/n \to 0$, $k/(np) \to \infty$ with $\sqrt{k} b_{n,k} \to M$ finite it follows that

- when $\mu \neq 0$
$$\frac{k}{h_0(\frac{k}{np})} \log R_k(p) \to -4z_{1-\alpha/2}\sigma\gamma\mu,$$
- when $\mu = 0$
$$\frac{\sqrt{k}}{b_{n,k}} \log R_k(p) \to -\frac{2z_{1-\alpha/2}\sigma}{\rho} > 0.$$

Concerning the relative width we can conclude that in case of the ridge estimator $\hat{Q}_k^{\mathrm{ridge}}(1-p)$ with $\mu > 0$ the new confidence interval (3.21) has a small asymptotic advantage over (3.16), while the opposite holds in all other cases using bias-reduced estimators. So in the case $\beta > 0$ we find that the confidence intervals (3.21) in combination with the ridge estimator are asymptotically advantageous. The case $\beta > 0$ corresponds to the situation where the Hill estimator has a positive bias, which is quite common in practice. See also limit relation (3.3) with $x > 1$ when $\ell_T(tx) > \ell_T(t)$ leading to $b(x) > 0$ and hence $\beta > 0$.

For the intervals to have coverage probabilities which are as close as possible to $1-\alpha$, the rate of convergence of $T_{k,n}^{\bullet}$ and $S_{k,n}^{\bullet}$ to normality is of major importance. In the simulations section we study the deviations from normality for sample sizes $n = 200$ and $n = 1000$ using normal QQ plots based on the standardized statistics $T_{k,n}^{\bullet}$ and $S_{k,n}^{\bullet}$. From this it appears that the imputation

Stellenbosch University https://scholar.sun.ac.za

of $\hat{\gamma}_k^\bullet$ in the denominator of $T_{k,n}^\bullet$ leads to clear deviations from normality, often leading to left skewness resulting from the right skewness of the distribution of $\hat{\gamma}_k^\bullet$. In case of the least squares estimator deviations become apparent both at the right and left tail. On the other hand, the distribution of the newly proposed pivotal quantity $S_{k,n}^\bullet$ is quite close to the distribution of $T_{k,n}^\bullet$ where the correct value of γ is substituted in the denominator.

The normal approximation can be improved by adding terms to its distribution function in order to explicitly correct for skewness, as is done when using Edgeworth expansions. An example of the application of Edgeworth expansions in extreme value theory is given by Haeusler and Segers (2007) who derived Edgeworth expansions for the standardized Hill estimator, and used them to derive asymptotic expansions for the coverage probabilities of a number of two-sided confidence intervals for the EVI. In case of the bias-reduced estimators, this could be pursued in future work. In the next section we propose saddlepoint approximations instead.

3.3 Saddlepoint confidence intervals

The saddlepoint method was developed in order to improve asymptotic normal approximations for small sample sizes. This method stems from complex analysis and was introduced in statistics by Daniels (1954). He applied the saddlepoint method to the Fourier inversion formula of the density function of a sample mean to yield an asymptotic estimate. In his paper he further showed that the saddlepoint approximation is equivalent to an optimal indirect application of the Edgeworth expansion, whereby the resulting asymptotic estimate has a uniform relative error over its entire domain which is of a lower order than either the normal approximation or the direct Edgeworth expansion. A further saddlepoint approximation was derived by Lugannani and Rice (1980) who applied the saddlepoint method of Bleistein (1966) to the Fourier inversion formula of the distribution function of a sample mean. Jensen (1995) generalized the saddlepoint approximations for the density and distribution functions of a sample mean to weighted means of independent random variables, among others. Butler (2007) gives a thorough overview of the saddlepoint approximation and its applications to univariate, multivariate and conditional random variables. See also section 6.5 in Jensen (1995) for a discussion on saddlepoint methods for independent non-identically distributed variables.

The saddlepoint approximation is based on the Fourier inversion formula and therefore requires that the characteristic function of the statistics to be known,

Stellenbosch University https://scholar.sun.ac.za

and then it yields asymptotic estimates for the density and distribution function, both of which are at least as accurate as the normal approximation or the Edgeworth expansion over the entire domain of the distribution. We now focus on deriving a saddlepoint approximation for the distribution function of $\frac{\log \hat{Q}_k^\bullet(1-p)-\log X_{n-k,n}}{\log Q(1-p)-\log X_{n-k,n}}$ by using the results of the following corollary. To this end we write $\hat{\gamma}_k^\bullet$ and $\hat{b}_{n,k}^\bullet$ as weighted means of the log-spacings Z_j:

$$\hat{\gamma}_k^\bullet = \frac{1}{k}\sum_{j=1}^{k} \hat{\mu}_{j,k}^\bullet Z_j \qquad \text{and} \qquad \hat{b}_{n,k}^\bullet = \frac{1}{k}\sum_{j=1}^{k} \hat{\nu}_{j,k}^\bullet \, Z_j,$$

where the weights $\hat{\mu}_{j,k}^\bullet$ and $\hat{\nu}_{j,k}^\bullet$ can be derived from Table 3.2 for the different estimators.

Corollary 3.2. *Assume $k, n \to \infty$ with $k/n \to 0$ such that $b_{n,k} = O(k^{-1/2})$ and that $p = p_n$ satisfies $k/(np_n) \to \infty$. Then we have under the second order condition (3.3) that*

$$\frac{S_{k,n}^\bullet}{\sqrt{k}} + 1 = \frac{1}{k}\sum_{j=1}^{k} \hat{\lambda}_{j,k}^\bullet \, \theta_{j,k} \, E_j + o_p\left(b_{n,k}\right) + o_p\left(\frac{1}{\sqrt{k}\, h_0(\frac{k}{np})}\right),$$

with $\hat{\lambda}_{j,k}^\bullet = \hat{\mu}_{j,k}^\bullet + \hat{\nu}_{j,k}^\bullet \kappa_{\hat{\rho}}\left(\frac{k}{np}\right)$ *and* $\theta_{j,k} = \frac{1+\beta\left(\frac{n}{k}\right)^{\rho} c_{j,k}}{1+\beta\left(\frac{n}{k}\right)^{\rho} \kappa_\rho\left(\frac{k}{np}\right)}$ *for* $j = 1, \ldots, k$.

Proof It follows from Theorem 1 that

$$\begin{aligned}
&\frac{S_{k,n}^\bullet}{\sqrt{k}} + 1 \\
=&\frac{\hat{\gamma}_k^\bullet + \hat{b}_{n,k}^\bullet \, \kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma + b_{n,k}\, \kappa_\rho\left(\frac{k}{np}\right)} \left(1 - \frac{\log(\frac{n}{k}\, U_{k+1,n})}{h_0(\frac{k}{np})}\right) + o_p\left(\frac{1}{\sqrt{k}\, h_0(\frac{k}{np})}\right) \\
=&\left(\frac{1}{k}\sum_{j=1}^{k} \frac{\hat{\mu}_{j,k}^\bullet + \hat{\nu}_{j,k}^\bullet \kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma + b_{n,k}\, \kappa_\rho\left(\frac{k}{np}\right)} Z_j\right) \left(1 + O_p\left(\frac{1}{\sqrt{k}\, h_0(\frac{k}{np})}\right)\right) + o_p\left(\frac{1}{\sqrt{k}\, h_0(\frac{k}{np})}\right) \\
=&\frac{1}{k}\sum_{j=1}^{k} \hat{\lambda}_{j,k}^\bullet \, \frac{1+\beta\,\left(\frac{k}{n}\right)^{-\rho} c_{j,k}}{1+\beta\,\left(\frac{k}{n}\right)^{-\rho} \kappa_\rho\left(\frac{k}{np}\right)} E_j \; + o_p\left(b_{n,k}\right) + o_p\left(\frac{1}{\sqrt{k}\, h_0(\frac{k}{np})}\right) \\
=&\frac{1}{k}\sum_{j=1}^{k} \hat{\lambda}_{j,k}^\bullet \, \theta_{j,k} \, E_j + o_p\left(b_{n,k}\right) + o_p\left(\frac{1}{\sqrt{k}\, h_0(\frac{k}{np})}\right).
\end{aligned}$$

Stellenbosch University https://scholar.sun.ac.za

Concerning the third step in the above derivation set

$$w_{j,k} = j\,\frac{\hat{\mu}^{\bullet}_{j,k} + \hat{\nu}^{\bullet}_{j,k}\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}{\gamma + b_{n,k}\,\kappa_{\rho}\left(\frac{k}{np}\right)}$$

with $\frac{1}{k}\sum_{j=1}^{k}\frac{w_{j,k}}{j}Z_j = \frac{1}{k}\sum_{j=1}^{k}\frac{w_{j,k}}{j}\,(\gamma + b_{n,k}c_{j,k})\,E_j + \frac{1}{k}\sum_{j=1}^{k} w_{j,k}\,\frac{R_{j,n}}{j}$, and

$$\begin{aligned}\frac{1}{k}\sum_{j=1}^{k} w_{j,k}\,\frac{R_{j,n}}{j} &= \frac{1}{k}\sum_{j=1}^{k}\left(\sum_{m=1}^{j}(w_{m,k} - w_{m-1,k})\right)\frac{R_{j,n}}{j}\\ &= \frac{1}{k}\sum_{m=1}^{k}(w_{m,k} - w_{m-1,k})\sum_{j=m}^{k}\frac{R_{j,n}}{j}\end{aligned}$$

(with $w_{0,k} := 0$) where $\sum_{j=m}^{k}\frac{R_{j,n}}{j} = o_p(b_{n,k})\log(k/m)$ as given in Theorem 4.1 in Beirlant et al. (2004). In case of the ridge estimators $w_{m,k} - w_{m-1,k} = C(m/k)^{-\hat{\rho}}$ and so $\frac{1}{k}\sum_{j=1}^{k} w_{j,k}\,\frac{R_{j,n}}{j} = o_p(b_{n,k})$.

Table 3.2: Parameters $\{\hat{\lambda}^{\bullet}_{j,k}\}$ corresponding to each quantile estimator

Quantile estimator $(\hat{Q}^{\bullet}_k)$	$\hat{\lambda}^{\bullet}_{j,k} = \hat{\mu}^{\bullet}_{j,k} + \hat{\nu}^{\bullet}_{j,k}\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)$
$\hat{Q}^{\mathrm{H}}_k$	1
$\hat{Q}^{\mathrm{LS}}_k$	$1 + \hat{\nu}^{\mathrm{LS}}_{j,k}\left(\kappa_{\hat{\rho}}\left(\frac{k}{np}\right) - \bar{d}_k\right)$
$\hat{Q}^{\mathrm{ridge}}_k$	$1 + \hat{\nu}^{\mathrm{ridge}}_{j,k}\left(\kappa_{\hat{\rho}}\left(\frac{k}{np}\right) - \bar{d}_k\right)$
$\hat{Q}^{\mathrm{CH}}_k$	$\left(1 - \frac{\hat{\beta}\left(\frac{n}{k}\right)^{\hat{\rho}}}{1-\hat{\rho}}\right)\left(1 + \hat{\beta}\left(\frac{n}{k}\right)^{\hat{\rho}}\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)\right)$

We now discuss the saddlepoint approximation of the distribution function of the approximation $\frac{1}{k}\sum_{j=1}^{n}\hat{\lambda}^{\bullet}_{j,k}\theta_{j,k}E_j$ which, from Corollary 3.2 itself is an approximation of $\frac{\log\hat{Q}^{\bullet}_k(1-p) - \log X_{n-k,n}}{\log Q(1-p) - \log X_{n-k,n}}$. The cumulant generating function of $\frac{1}{k}\sum_{j=1}^{n}\hat{\lambda}^{\bullet}_{j,k}\theta_{j,k}E_j$ is given by

$$K_k(z;\underline{\lambda}_k) = \log\mathbb{E}\left[e^{z\,\frac{1}{k}\sum_{j=1}^{k}\hat{\lambda}_{j,k}\theta_{j,k}E_j}\right] = -\sum_{j=1}^{k}\log\left(1 - \frac{\hat{\lambda}_{j,k}\theta_{j,k}}{k}\,z\right).$$

Stellenbosch University https://scholar.sun.ac.za

The saddlepoint approximation of the distribution function of $\frac{1}{k}\sum_{j=1}^{n}\hat{\lambda}_{j,k}\theta_{j,k}E_j$ is readily derived from Butler (2007) yielding the following asymptotic estimate:

$$F\left(x;\lambda\right)=\Phi(w)+\phi(w)\left\{\frac{1}{w}-\frac{1}{\psi\sqrt{K_k''(\psi)}}\right\}, \tag{3.22}$$

where ϕ and Φ are the standard normal density and distribution function respectively, and ψ is the unique real solution to the defining equation

$$K_k'(\psi;\underline{\lambda}_k)=x \tag{3.23}$$

and $w=\text{sign}\{\psi\}\sqrt{2}\sqrt{\psi\,x-K_k(\psi;\underline{\lambda}_k)}$ (see Bleistein (1966)).

The parameters $\{\lambda_{j,k}\}$ corresponding to each quantile estimator are summarized in Table 3.2. In practice, the parameters $\theta_{j,k}$ are of course estimated by $\hat{\theta}_{j,k}=\frac{1+\hat{\beta}\left(\frac{n}{k}\right)^{\hat{\rho}}d_{j,k}}{1+\hat{\beta}\left(\frac{n}{k}\right)^{\hat{\rho}}\kappa_{\hat{\rho}}\left(\frac{k}{np}\right)}$.

The saddlepoint confidence interval for $Q(1-p)$ with a $(1-\alpha)100\%$ level of significance is given by

$$\left[X_{n-k,n}\,e^{\frac{\hat{\gamma}_k^{\bullet}\,h_0\left(\frac{k}{np}\right)+b_{n,k}^{\bullet}\,h_{\hat{\rho}}\left(\frac{k}{np}\right)}{q_F\left(1-\frac{\alpha}{2};\underline{\lambda}_k\right)}},\;X_{n-k,n}\,e^{\frac{\hat{\gamma}_k^{\bullet}\,h_0\left(\frac{k}{np}\right)+b_{n,k}^{\bullet}\,h_{\hat{\rho}}\left(\frac{k}{np}\right)}{q_F\left(\frac{\alpha}{2};\underline{\lambda}_k\right)}}\right] \tag{3.24}$$

where $q_F(\beta;\underline{\lambda}_k)$ denotes the β quantile of the saddlepoint approximations (3.22):

$$q_F\left(\alpha;\underline{\lambda}_k\right)=\{x:F\left(x;\underline{\lambda}_k\right)=\alpha\}$$

The saddlepoint approximation applied to the weighted mean $\frac{1}{k}\sum_{j=1}^{k}\hat{\lambda}_{j,k}^{\bullet}\,\theta_{j,k}\,E_j$ has a much lower relative error than the normal approximation, and furthermore, this lower relative error holds uniformly over the entire domain of the distribution, while the normal approximation's relative error only holds in the central domain of the distribution. This implies that the saddlepoint approximation is more accurate than the normal approximation, see Proposition 6.1.1 in Jensen (1995).

In the simulation section below we study the accuracy of the saddlepoint approximation of $1+S_{k,n}^{\bullet}/\sqrt{k}=\frac{1}{k}\sum_{j=1}^{k}\hat{\lambda}_{j,k}^{\bullet}\,\theta_{j,k}\,E_j$ as an approximation of the distribution of $\frac{\log\hat{Q}_k^{\bullet}(1-p)-\log X_{n-k,n}}{\log Q(1-p)-\log X_{n-k,n}}$ for sample sizes of $n=200$. First we produced a QQ plot of the distribution of $1+S_{k,n}^{\bullet}/\sqrt{k}$ against the saddlepoint distribution of $\frac{1}{k}\sum_{j=1}^{k}\lambda_{j,k}^{\bullet}\,\theta_{j,k}\,E_j$, i.e. with the correct values of

Stellenbosch University https://scholar.sun.ac.za

the different parameters in the saddlepoint distribution. Next for each generated sample we study the deviations between the saddlepoint approximations of $\frac{1}{k}\sum_{j=1}^{k}\hat{\lambda}_{j,k}^{\bullet}\,\hat{\theta}_{j,k}\,E_j$ versus the saddlepoint approximation of $\frac{1}{k}\sum_{j=1}^{k}\lambda_{j,k}^{\bullet}\,\theta_{j,k}\,E_j$ which shows that the error propagated by the imputation of second order parameters is much smaller than the error produced by using the approximation $\frac{1}{k}\sum_{j=1}^{k}\lambda_{j,k}^{\bullet}\,\theta_{j,k}\,E_j$ for $1+S_{k,n}^{\bullet}/\sqrt{k}$.

3.4 Simulation study

In this section the characteristics of the confidence intervals of large quantiles are illustrated for a variety of different distributions with a simulation study, taking 10 000 repetitions of samples of size $n = 200$ and $n = 1000$. The following distributions are used:

- *The Fréchet*(2) *distribution* with $\bar{F}(x) = 1 - \exp\left(-x^{-2}\right)$ for $x > 0$, so that $\gamma = \frac{1}{2}$, $\rho = -1$ and $\beta = \frac{1}{2}$.
- *The Burr*$\left(\sqrt{2},\sqrt{2}\right)$ *distribution* with $\bar{F}(x) = \left(1 + x^{\sqrt{2}}\right)^{-\sqrt{2}}$ for $x > 0$, so that $\gamma = \frac{1}{2}$, $\rho = -\frac{\sqrt{2}}{2}$ and $\beta = 1$.
- *The loggamma(2,2) distribution* with $\bar{F}(x) = x^{-2}\,(1 + 2\log x)$ for $x > 1$, so that $\gamma = \frac{1}{2}$ and $\rho = 0$. Note that this case falls outside the scope of the mathematical conditions for the theoretical results where we assumed $\rho < 0$.

The normal approximation of $S_{50,200}^{\bullet}$ from Corollary 3.1 is illustrated in Figure 3.1, jointly with the normal approximation of $T_{50,200}^{\bullet}$ from (3.15) with γ in the denominator of $T_{50,200}^{\bullet}$ being equal to the correct value next to being estimated with the corresponding $\hat{\gamma}_{50,200}^{\bullet}$. It follows that the normal approximation for $S_{k,n}^{\bullet}$ and $T_{k,n}^{\bullet}$ with γ equal to the correct value do equally well. The differences from the 45 degree line are induced by the fact that the limiting normal parameters (μ,σ) are different from $(0,1)$. Imputation of $\hat{\gamma}^{\bullet}$ in $T_{50,200}^{\bullet}$ does often induces left skewness in the distribution of $T_{50,200}^{\bullet}$. Note that in case of the Fréchet and loggamma distribution the distribution of $T_{50,200}^{\mathrm{LS}}$ shows heavy tails on both sides for all three estimation methods.

The saddlepoint approximation of $1 + S_{50,200}^{\bullet}/\sqrt{50}$ is illustrated in Figures 3.2 and 3.3: first we show QQ plots of the distribution of $1 + S_{50,200}^{\bullet}/\sqrt{50}$ against the saddlepoint distribution of $\frac{1}{50}\sum_{j=1}^{50}\lambda_{j,50}^{\bullet}\,\theta_{j,50}\,E_j$, i.e. with the

Stellenbosch University https://scholar.sun.ac.za

correct values of the different parameters in the saddlepoint distribution. From this only a shift between the two distributions appears. Next for each generated sample we study the deviations between the saddlepoint approximations of $\frac{1}{50}\sum_{j=1}^{50}\hat{\lambda}^{\bullet}_{j,50}\,\hat{\theta}_{j,50}\,E_j$ versus the saddlepoint approximation of $\frac{1}{50}\sum_{j=1}^{50}\lambda^{\bullet}_{j,50}\,\theta_{j,50}\,E_j$ which shows that the error propagated by the imputation of second order parameters is much smaller than the error produced by using the approximation $\frac{1}{50}\sum_{j=1}^{50}\lambda^{\bullet}_{j,50}\,\theta_{j,50}\,E_j$ for $1+S^{\bullet}_{50,200}/\sqrt{50}$.

In Figures 3.4 to 3.9 we illustrate the utility of each of the new confidence intervals proposed in sections 3.2 and 3.3 based on $\hat{Q}^{\mathrm{H}}_k$, $\hat{Q}^{\mathrm{LS}}_k$, $\hat{Q}^{\mathrm{ridge}}_k$ and $\hat{Q}^{\mathrm{CH}}_k$ compared to the classical confidence intervals discussed in section 3.2 for small ($n=200$) and large ($n=1000$) sample sizes. A comparison is proposed of the coverage proportion, the relative bias of the corresponding quantile estimator using

$$RB^{\bullet}_{k,n}=\frac{\hat{Q}^{\bullet}_{k,n}(p)}{Q(p)}-1,$$

and the width of the two-sided confidence intervals using the relative measure

$$RW^{\bullet}_{k,n}=\frac{U^{\bullet}_k(p)}{L^{\bullet}_k(p)}$$

where $U^{\bullet}_k(p)$ and $L^{\bullet}_k(p)$ are the upper and lower confidence bounds given by either confidence interval (3.16), (3.21), and (3.24).

The following conclusions can be drawn:

- Concerning the different quantile estimators, $\hat{Q}^{\mathrm{CH}}_k$ and $\hat{Q}^{\mathrm{ridge}}_k$ perform much more accurate than the $\hat{Q}^{\mathrm{H}}_k$, both in terms of coverage probabilities and interval width for a longer interval of k values. $\hat{Q}^{\mathrm{ridge}}_k$ offers a good compromise between the three bias reduced methods. In some cases such as the Burr and log-gamma, the intervals based on $\hat{Q}^{\mathrm{LS}}_k$ are much wider than those based on $\hat{Q}^{\mathrm{ridge}}_k$ and $\hat{Q}^{\mathrm{CH}}_k$. Furthermore in several instances the width of the $\hat{Q}^{\mathrm{LS}}_k$-based intervals is quite unstable or very large.
- Concerning the three approximation methods given in (3.15) and Corollaries 3.1 and 3.2, the saddlepoint confidence intervals are narrow over a large set of k values, while the normal approximation based on pivot $S^{\bullet}_{k,n}$ often shows accurate coverage probabilities over a longer interval of k values especially in case of the ridge regression estimator.

Stellenbosch University https://scholar.sun.ac.za

- The above conclusions mainly hold for small sample sizes. In case $n = 1000$ the observed differences become less apparent.

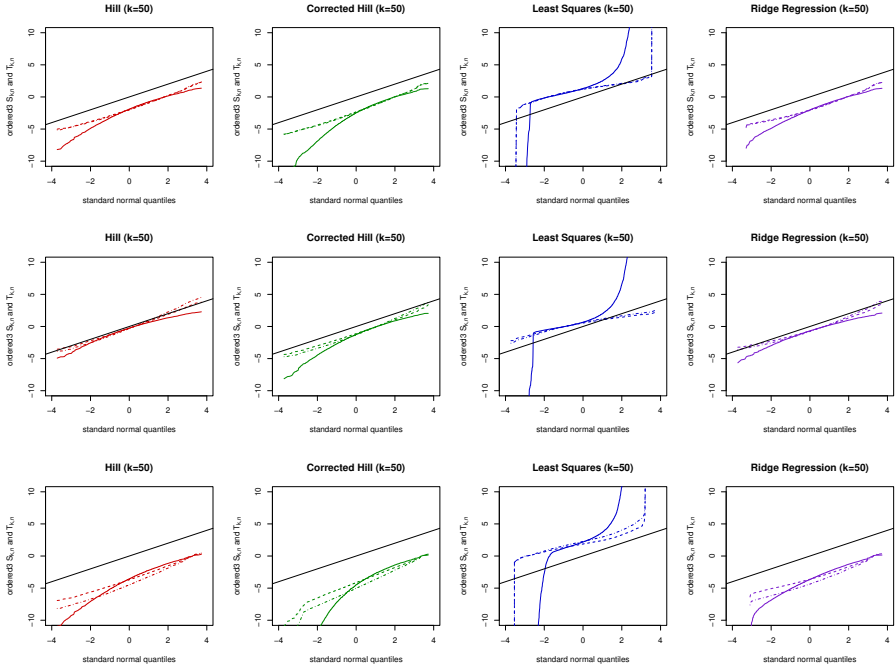

Figure 3.1: Normal QQ plots of 1000 simulated $S^{\bullet}_{50,200}$ values (dashed) and $T^{\bullet}_{50,200}$ values with γ estimated (solid) and with the actual value of γ (dash dotted): Fréchet (top), Burr (center) and loggamma (bottom) distribution; from left to right $\hat{Q}^{\mathrm{H}}_k(.995)$, $\hat{Q}^{\mathrm{CH}}_k(.995)$, $\hat{Q}^{\mathrm{LS}}_k(.995)$ and $\hat{Q}^{\mathrm{ridge}}_k(.995)$.

Stellenbosch University https://scholar.sun.ac.za

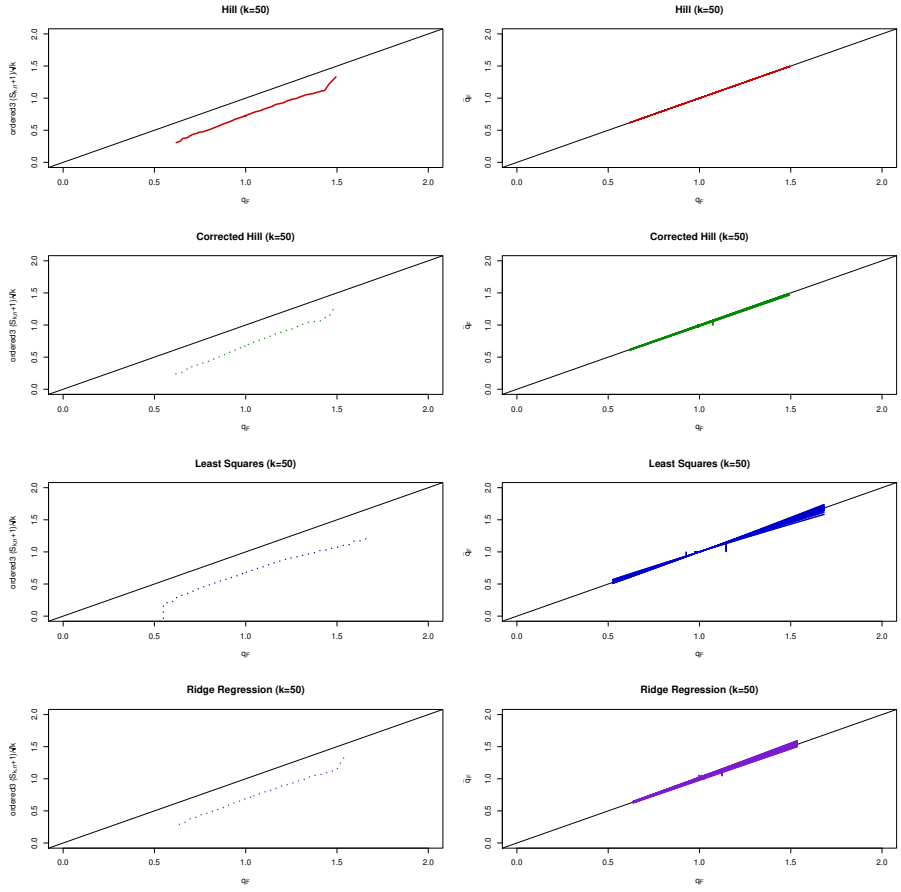

Figure 3.2: Fréchet distribution with $\gamma = \frac{1}{2}$ and $\rho = -1$: QQ plots of 1000 simulated $1 + S^{\bullet}_{50,200}/\sqrt{50}$ values against the saddlepoint distribution with correct parameters (left); QQ plots of $j/101$ $(j = 1, \ldots, 100)$ quantiles of the saddlepoint distribution with correct parameters against 1000 saddlepoint distributions with estimated parameters (right). From top to bottom $\hat{Q}^{\mathrm{H}}_k(.995)$, $\hat{Q}^{\mathrm{CH}}_k(.995)$, $\hat{Q}^{\mathrm{LS}}_k(.995)$ and $\hat{Q}^{\mathrm{ridge}}_k(.995)$.

Stellenbosch University https://scholar.sun.ac.za

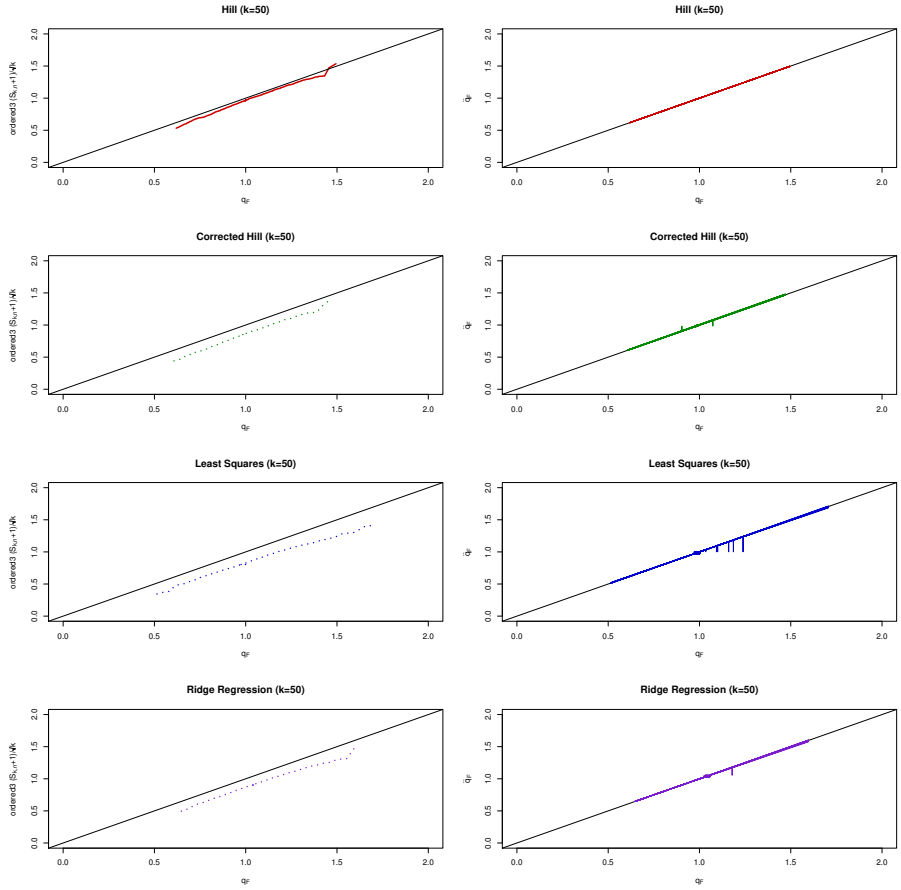

Figure 3.3: Burr distribution with $\gamma = \frac{1}{2}$ and $\rho = -\frac{\sqrt{2}}{2}$: QQ plots of 1000 simulated $1 + S^{\bullet}_{50,200}/\sqrt{50}$ values against the saddlepoint distribution with correct parameters (left); QQ plots of $j/101$ $(j = 1, \ldots, 100)$ quantiles of the saddlepoint distribution with correct parameters against the 1000 saddlepoint distributions with estimated parameters (right). From top to bottom $\hat{Q}_k^{\mathrm{H}}(.995)$, $\hat{Q}_k^{\mathrm{CH}}(.995)$, $\hat{Q}_k^{\mathrm{LS}}(.995)$ and $\hat{Q}_k^{\mathrm{ridge}}(.995)$.

Stellenbosch University https://scholar.sun.ac.za

3.4.1 Fréchet distribution

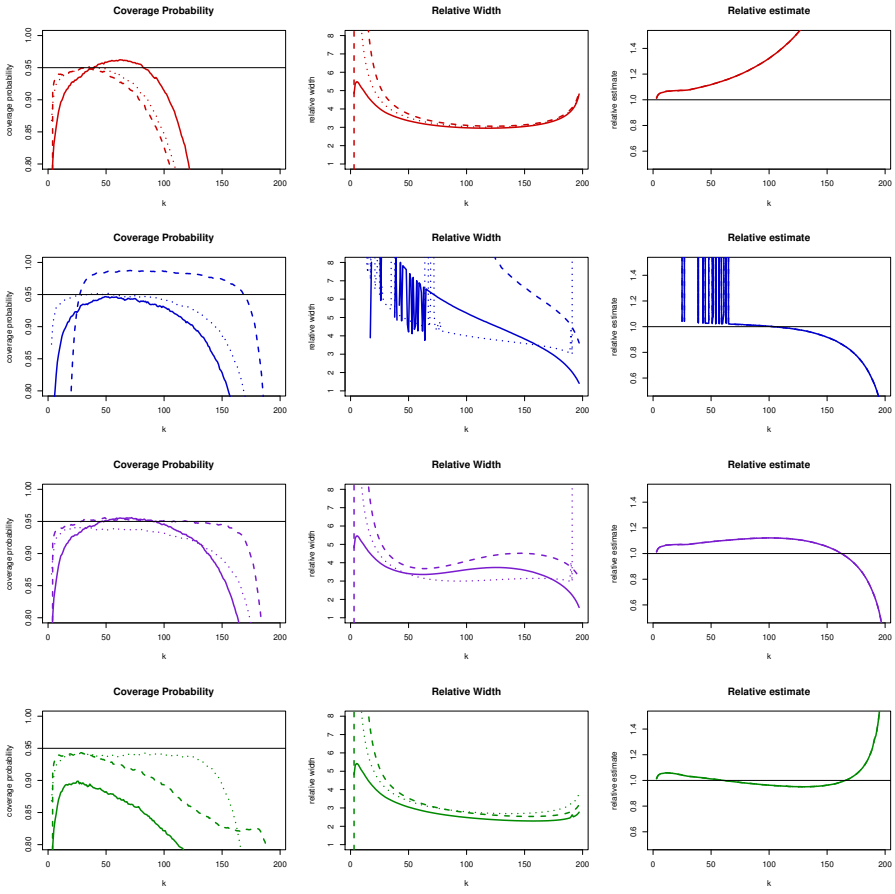

Figure 3.4: The coverage probabilities (left), widths (center) and estimator relative bias (right) of the 95% confidence intervals for the .995 quantile of the Fréchet distribution with $\gamma = \frac{1}{2}$ and $\rho = -1$, sample size $n = 200$. These confidence intervals are based on $\hat{Q}_k^{\mathrm{H}}(.995)$, $\hat{Q}_k^{\mathrm{LS}}(.995)$, $\hat{Q}_k^{\mathrm{ridge}}(.995)$ and $\hat{Q}_k^{\mathrm{CH}}(.995)$ (in that order from top to bottom) together with their classical normal (solid line), new normal (dashed line) and saddlepoint (dotted line) approximations.

Stellenbosch University https://scholar.sun.ac.za

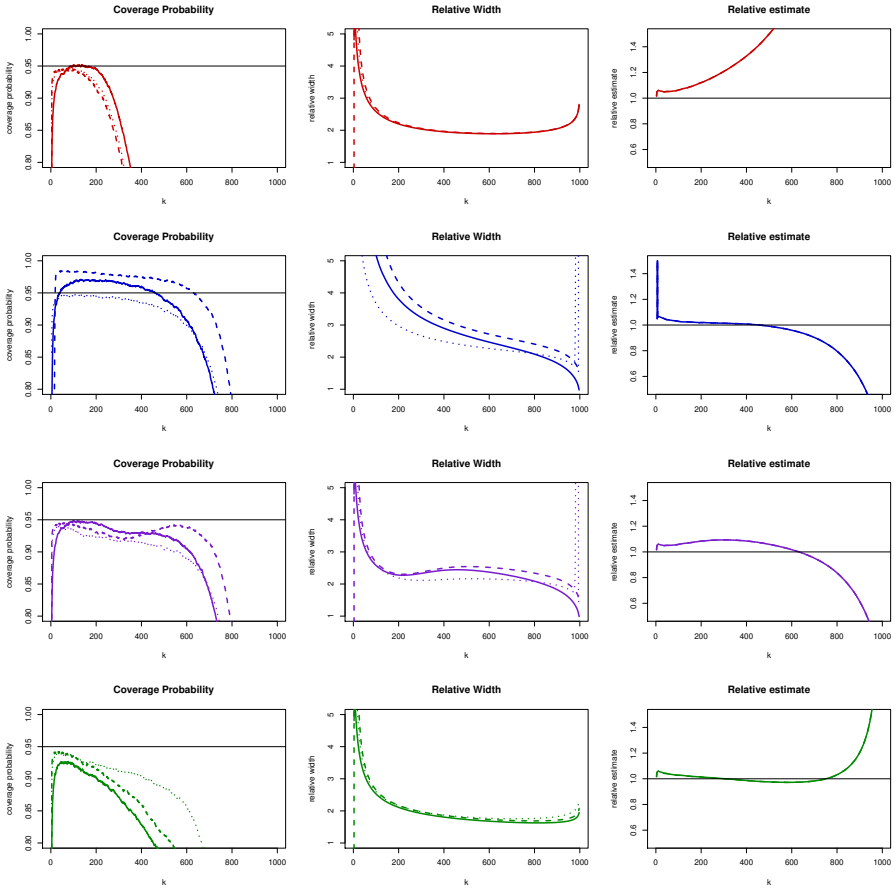

Figure 3.5: The coverage probabilities (left), widths (center) and estimator relative bias (right) of the 95% confidence intervals for the .999 quantile of the Fréchet distribution with $\gamma = \frac{1}{2}$ and $\rho = -1$, sample size $n = 1000$. These confidence intervals are based on $\hat{Q}_k^{\mathrm{H}}(.999)$, $\hat{Q}_k^{\mathrm{LS}}(.999)$, $\hat{Q}_k^{\mathrm{ridge}}(.999)$ and $\hat{Q}_k^{\mathrm{CH}}(.999)$ (in that order from top to bottom) together with their classical normal (solid line), new normal (dashed line) and saddlepoint (dotted line) approximations.

Stellenbosch University https://scholar.sun.ac.za

3.4.2 Burr distribution

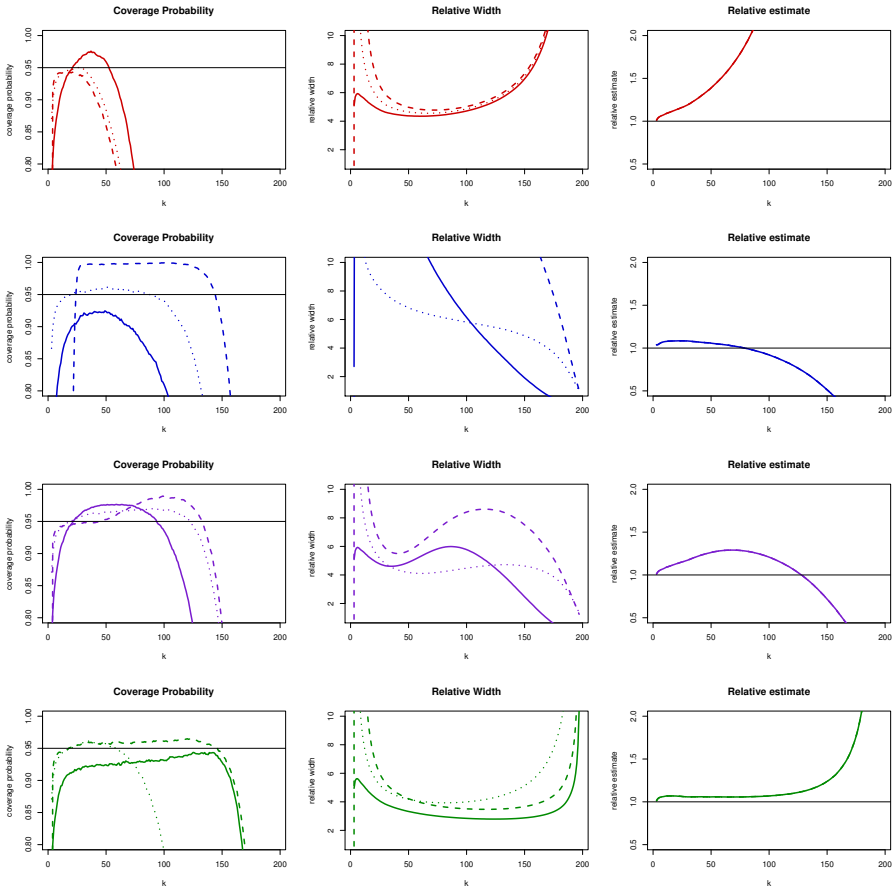

Figure 3.6: The coverage probabilities (left), widths (center) and estimator relative bias (right) of the 95% confidence intervals for the .995 quantile of the Burr distribution with $\gamma = \frac{1}{2}$ and $\rho = -\frac{\sqrt{2}}{2}$, sample size $n = 200$. These confidence intervals are based on $\hat{Q}_k^{\mathrm{H}}(.995)$, $\hat{Q}_k^{\mathrm{LS}}(.995)$, $\hat{Q}_k^{\mathrm{ridge}}(.995)$ and $\hat{Q}_k^{\mathrm{CH}}(.995)$ (in that order from top to bottom) together with their corresponding classical normal (solid line), new normal (dashed line) and saddlepoint (dotted line) approximations.

Stellenbosch University https://scholar.sun.ac.za

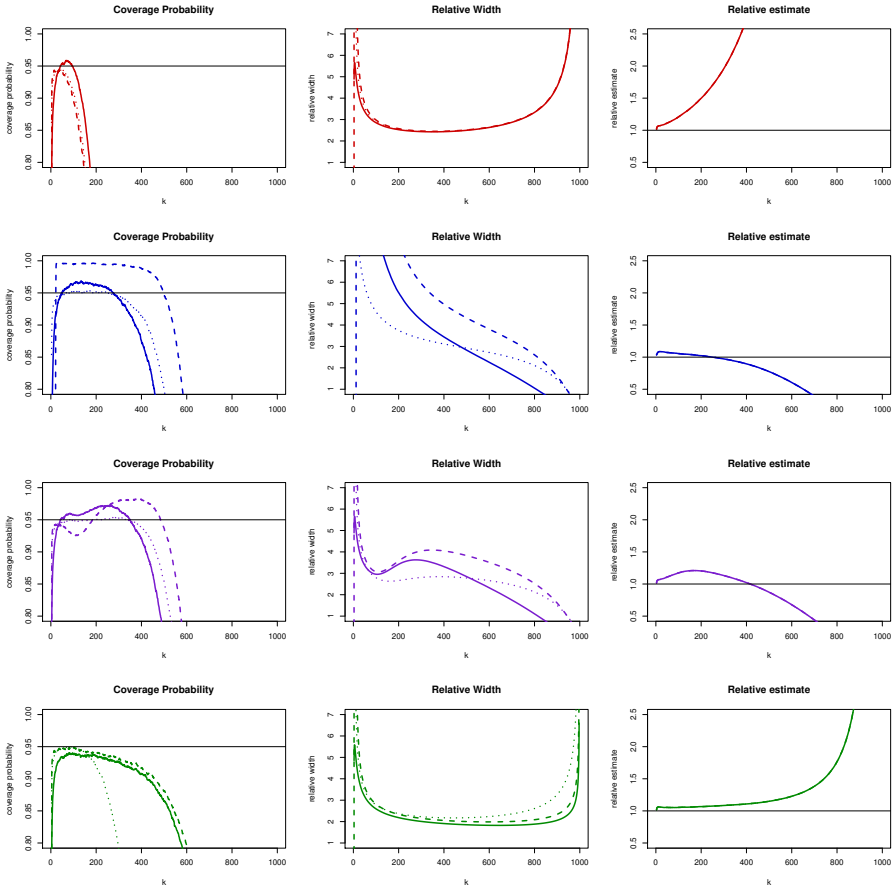

Figure 3.7: The coverage probabilities (left), widths (center) and estimator relative bias (right) of the 95% confidence intervals for the .995 quantile of the Burr distribution with $\gamma = \frac{1}{2}$ and $\rho = -\frac{\sqrt{2}}{2}$, sample size $n = 1000$. These confidence intervals are based on $\hat{Q}_k^{\mathrm{H}}(.999)$, $\hat{Q}_k^{\mathrm{LS}}(.999)$, $\hat{Q}_k^{\mathrm{ridge}}(.999)$ and $\hat{Q}_k^{\mathrm{CH}}(.999)$ (in that order from top to bottom) together with their corresponding classical normal (solid line), new normal (dashed line) and saddlepoint (dotted line) approximations.

Stellenbosch University https://scholar.sun.ac.za

3.4.3 Log Gamma distribution

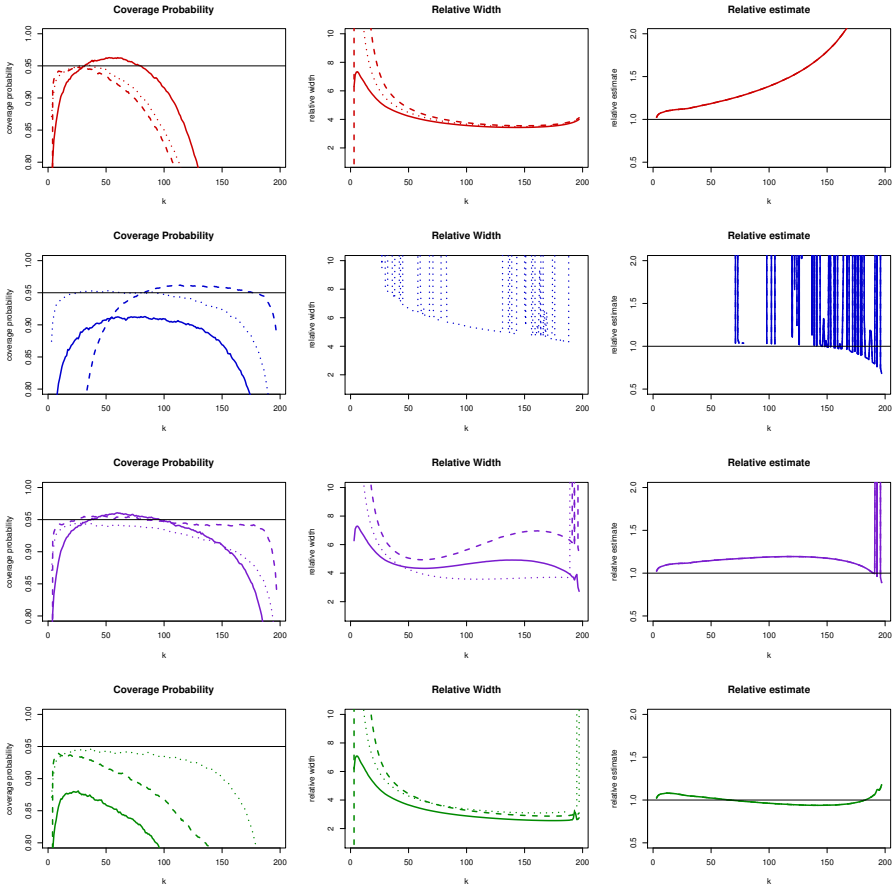

Figure 3.8: The coverage probabilities (left), widths (center) and estimator relative bias (right) of the 95% confidence intervals for the .995 quantile of the loggamma distribution with $\gamma = \frac{1}{2}$ and $\rho = 0$, sample size $n = 200$. These confidence intervals are based on $\hat{Q}_k^{\mathrm{H}}(.995)$, $\hat{Q}_k^{\mathrm{LS}}(.995)$, $\hat{Q}_k^{\mathrm{ridge}}(.995)$ and $\hat{Q}_k^{\mathrm{CH}}(.995)$ (in that order from top to bottom) together with their corresponding classical normal (solid line), new normal (dashed line) and saddlepoint (dotted line) approximations.

Stellenbosch University https://scholar.sun.ac.za

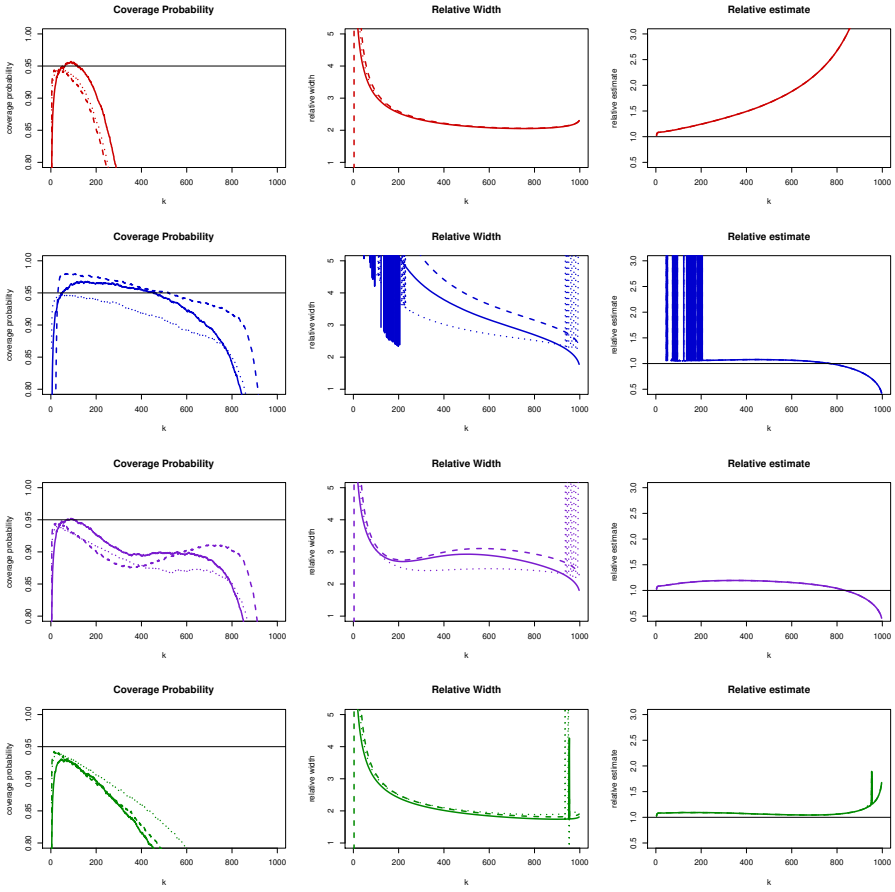

Figure 3.9: The coverage probabilities (left), widths (center) and estimator relative bias (right) of the 95% confidence intervals for the .999 quantile of the loggamma distribution with $\gamma = \frac{1}{2}$ and $\rho = 0$, sample size $n = 1000$. These confidence intervals are based on $\hat{Q}_k^{\mathrm{H}}(.999)$, $\hat{Q}_k^{\mathrm{LS}}(.999)$, $\hat{Q}_k^{\mathrm{ridge}}(.999)$ and $\hat{Q}_k^{\mathrm{CH}}(.999)$ (in that order from top to bottom) together with their corresponding classical normal (solid line), new normal (dashed line) and saddlepoint (dotted line) approximations.

Stellenbosch University https://scholar.sun.ac.za

3.5 Case study: Secura Re data

The methods outlined in this paper are used to construct confidence intervals for the once-in-50-years and once-in-100-years Secura Re claim size based on the data set from Beirlant et al. (2004) which consists of 371 observed claim sizes during the years 1988 to 2001. It is assumed that claims occur uniformly over each year which implies that the $.9993^{\text{th}}$ and $.99965^{\text{th}}$ quantile corresponds to the once-in-50-years and once-in-100-years claim size.

This example clearly illustrates the advantage of using the ridge regression quantile estimator which shows a strong reduction in the width of its corresponding confidence interval for small values of k, while the simulation study did indicate that the coverage probability is quite reliable.

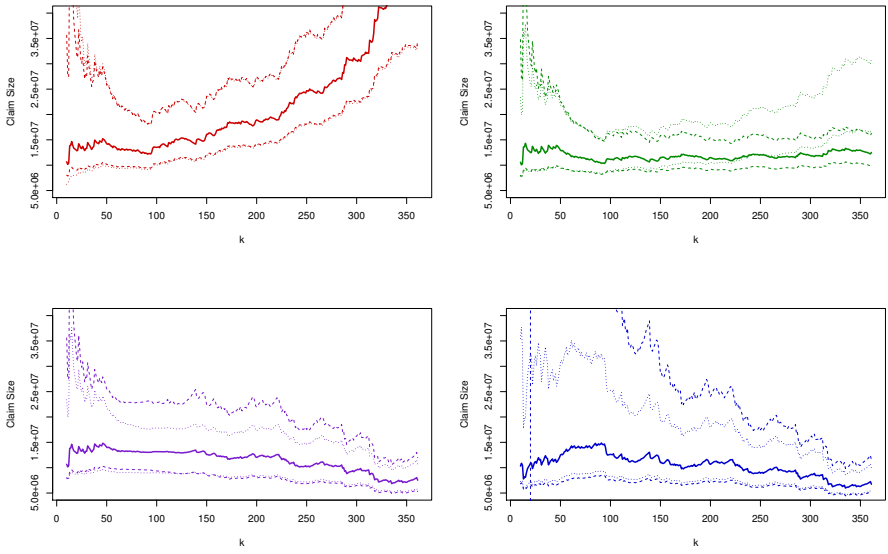

Figure 3.10: Secura Re data: the once-in-50-years claim size estimates and their corresponding 95% upper confidence bounds based on $\hat{Q}_k^{\text{H}}$ top left), $\hat{Q}_k^{\text{CH}}$ (top right), $\hat{Q}_k^{\text{ridge}}$ (bottom left) and $\hat{Q}_k^{\text{LS}}$ (bottom right). The thick solid line represents the quantile estimate, and the thin dashed line, respectively the thin solid line, represent the saddlepoint respectively the normal upper confidence bounds.

Stellenbosch University https://scholar.sun.ac.za

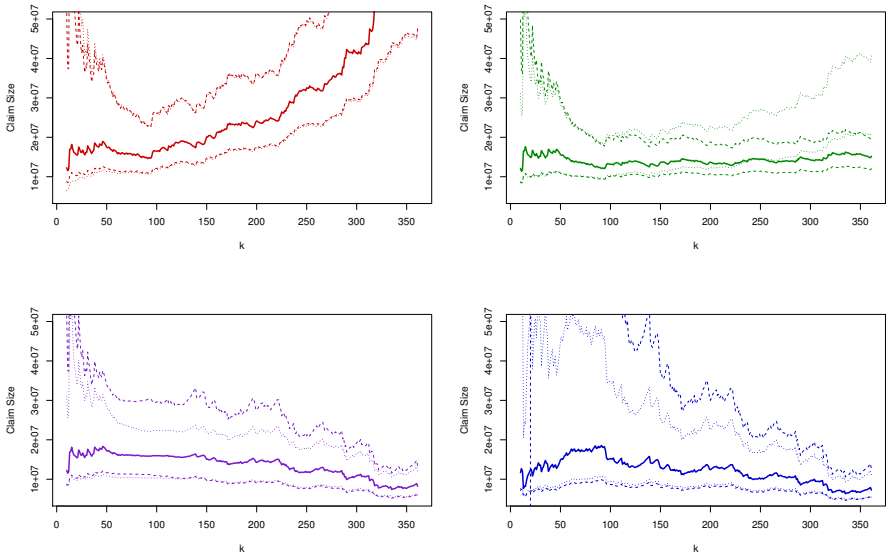

Figure 3.11: Secura Re data: the once-in-100-years claim size estimates and their corresponding 95% upper confidence bounds based on $\hat{Q}_k^{\mathrm{H}}$ (top left), $\hat{Q}_k^{\mathrm{CH}}$ (top right), $\hat{Q}_k^{\mathrm{ridge}}$ (bottom left) and $\hat{Q}_k^{\mathrm{LS}}$ (bottom right). The thick solid line represents the quantile estimate, and the thin dashed line, respectively the thin solid line, represent the saddlepoint respectively the normal upper confidence bounds.

3.6 Conclusion

The aim of this paper is to construct accurate and narrow confidence intervals for extreme quantiles of Pareto-type distributions.

We consider three different extreme quantile bias reduced quantile estimators. A novel standardization of these quantile estimators is introduced together with its corresponding asymptotic normal and saddlepoint approximations, which then are used to construct new asymptotic normal and saddlepoint confidence intervals for extreme quantiles.

These confidence intervals are assessed in terms of coverage probability and interval width by means of small sample simulations and a case study, with

Stellenbosch University https://scholar.sun.ac.za

special emphasis on small sample sizes. We also considered the region of k values where the coverage and width of the confidence intervals are acceptable. The ridge regression estimation method together with the saddlepoint approximation method generally provides accurate confidence intervals both concerning coverage accuracy and interval width.

Stellenbosch University https://scholar.sun.ac.za

Chapter 4

Empirical Center-Outward Quantiles: Applications to Risk Measurement

In Hallin (2017) the concepts of the empirical center-outward distribution and quantile functions in $\mathbb{R}^d$ were developed as generalizations of the classical univariate empirical distribution and quantile functions to higher dimensions. This approach is rooted in optimal transport theory and the center-outward distribution function is a continuous bijective cyclically monotone mapping from $\mathbb{R}^d$ to the unit ball $\mathbb{S}_d$. Here we propose a smooth interpolated version of the center-outward distribution function as an alternative to del Barrio et al. (2018), which allows for the computation of some new empirical risk measures. First, empirical risk measures can be based on the convex potential of the proposed mapping, while the volumes of the space inside the quantile contours lead to a second set of risk measures. We also discuss application of the transport to the evaluation of risk for multivariate regularly varying distributions. Some simulations and applications to case studies illustrate the value of these proposals.

4.1 Introduction

In order to circumvent the lack of left-to-right ordering in dimension d higher than one Hallin (2017) introduced a center-outward distribution function $\mathbf{F}_{\pm}$

Stellenbosch University https://scholar.sun.ac.za

of a non-vanishing Lebesgue-absolutely continuous distribution function of a random vector $\mathbf{X}$ as the unique gradient of a convex function which pushes the $\mathbf{X}$ distribution forward to the uniform distribution $\mathbf{U}_d$ over the unit ball $\mathbb{S}_d$ in $\mathbb{R}^d$, as defined in optimal transport theory described for instance in Villani (2009). Uniformity here refers to the product measure of a uniform distribution over the directions (the unit sphere $\mathbb{S}_{d-1}$) and a uniform distribution over the distances to the origin (the unit interval [0,1]). Of course, in case where data are all positive, i.e. $\mathbf{X} > \mathbf{0}$, then the transport can only be carried out towards the corresponding subsector of the unit sphere.

This extends the one-dimensional transformation $2F - 1$ from the domain of a random variable X to [-1,1], the unit sphere in $\mathbb{R}$. Boldface notation is used here in order to emphasize vector-valued objects. Clearly, $\mathbf{F}$ and $\mathbf{F}_{\pm}$ carry the same information about the $\mathbf{X}$ distribution, which they characterize. The inverse mapping $\mathbf{T}_{\pm}$ has the interpretation of a *center-outward quantile function*, mapping $\tau\mathbb{S}_{d-1}$ (the sphere with radius $\tau \in (0, 1]$) to the *center-outward quantile contour* $\mathbf{T}_{\pm}(\tau\mathbb{S}_{d-1})$ with probability content τ.

An interesting example is provided by the elliptical multivariate distributions. Given a location vector $\boldsymbol{\mu} \in \mathbb{R}^d$ and a positive-definite symmetric matrix Σ, $\mathbf{X}$ has elliptical distribution if and only if $\mathbf{Z} := \Sigma^{-1/2}(\mathbf{X} - \boldsymbol{\mu})$ has a spherical distribution, which holds if and only if

$$\mathbf{F}_{ell}(\mathbf{Z}) := \frac{\mathbf{Z}}{\|\mathbf{Z}\|} F_R(\|\mathbf{Z}\|) \sim \mathbf{U}_d, \tag{4.1}$$

with F_R the distribution function of $\|\mathbf{Z}\|$. In fact, Chernozhukov et al. (2017) showed that $\mathbf{F}_{ell}$ corresponds to $\mathbf{F}_{\pm}$. For the quantile version we then have

$$\mathbf{T}_{ell}(\mathbf{U}) := \frac{\mathbf{U}}{\|\mathbf{U}\|} Q_R(\|\mathbf{U}\|) \sim \mathbf{Z}, \tag{4.2}$$

with Q_R the quantile function of F_R.

The empirical counterparts $\mathbf{F}_{\pm}^{(n)}$ and $\mathbf{T}_{\pm}^{(n)}$ to $\mathbf{F}_{\pm}$ and $\mathbf{T}_{\pm}$ can be obtained as a cyclically monotone (discrete) mapping from the random sample $\mathbf{X}_1, \ldots, \mathbf{X}_n$ to a "uniform" grid over $\mathbb{S}_d$.

Writing $\mathbf{U}_{d,n}$ for the empirical measure on the points in the uniform grid, $\mathbf{u}_1, \ldots, \mathbf{u}_n$ and $\mathbf{U}_d$ the product of the uniform measure over the unit sphere, we assume that the choice of this grid ensures

$$\mathbf{U}_{d,n} \to_w \mathbf{U}_d, \tag{4.3}$$

Stellenbosch University https://scholar.sun.ac.za

with $\to_w$ denoting weak convergence of probability measures, as the direction is chosen uniformly over the unit sphere, the modulus is independent of the direction and the marginal distribution of the modulus converges to the uniform. In order to construct a uniform grid, in Hallin (2017) n_S points are chosen at random over the unit sphere and for each radius $r \in \{\frac{1}{n_R+1}, \ldots, \frac{n_R}{n_R+2}\}$ the n_S points in the unit sphere are scaled up to radius r. Additionally, some n_0 copies of $\mathbf{0}$ are added to complete a uniform grid of size $n = n_S n_R + n_0$. Here, in order to be able to obtain almost surely different quantile contours at the τ levels $\frac{i}{n+1}$ $(i = 1, \ldots, n)$, we generate grids $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ which combine these radii $r_i = \frac{i}{n+1}$ levels with a set of uniformly distributed points on the unit sphere $\mathcal{S}_{d-1}$. The construction of such grids $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$ is discussed further in section 4.2.

A subset $S = \{(\mathbf{x}_1, \mathbf{u}_1), \ldots, (\mathbf{x}_n, \mathbf{u}_n)\}$ of $\mathbb{R}^d \times \mathbb{R}^d$ is cyclically monotone if and only if

$$\langle \mathbf{u}_1, \mathbf{x}_2 - \mathbf{x}_1 \rangle + \langle \mathbf{u}_2, \mathbf{x}_3 - \mathbf{x}_2 \rangle + \ldots + \langle \mathbf{u}_n, \mathbf{x}_1 - \mathbf{x}_n \rangle \leq 0,$$

or, equivalently, if and only if S maximizes the empirical correlation $\sum_{i=1}^n \langle \mathbf{x}_i, \mathbf{u}_i \rangle$, or minimizes $\sum_{i=1}^n \|\mathbf{u}_i - \mathbf{x}_i\|^2$, where $\|\mathbf{x}\|$ stands for the Euclidean norm of $\mathbf{x} \in \mathbb{R}^d$.
Minimum least squares coupling $\mathbf{x}_1, \ldots, \mathbf{x}_n$ to the points $\mathbf{u}_1, \ldots, \mathbf{u}_n$ is equivalent to finding the entries in the cost matrix

$$C = \begin{bmatrix} \|\mathbf{u}_1 - \mathbf{x}_1\|^2 & \|\mathbf{u}_1 - \mathbf{x}_2\|^2 & \ldots & \|\mathbf{u}_1 - \mathbf{x}_n\|^2 \\ \|\mathbf{u}_2 - \mathbf{x}_1\|^2 & \|\mathbf{u}_2 - \mathbf{x}_2\|^2 & \ldots & \|\mathbf{u}_2 - \mathbf{x}_n\|^2 \\ \vdots & \vdots & \ddots & \vdots \\ \|\mathbf{u}_n - \mathbf{x}_1\|^2 & \|\mathbf{u}_n - \mathbf{x}_2\|^2 & \ldots & \|\mathbf{u}_n - \mathbf{x}_n\|^2 \end{bmatrix}$$

which lead to the smallest sum of entries. It is well known that the subdifferential of a convex function Ψ from $\mathbb{R}^d$ to $\mathbb{R}$ enjoys cyclical monotonicity, while the converse is also true as shown by Rockaffelar (1996): any cyclical monotone subset of $\mathbb{R}^d \times \mathbb{R}^d$ is contained in the subdifferential of some convex function.

The empirical center-outward distribution function $\mathbf{F}_{\pm}^{(n)}$ is only defined on the observed data through

$$\mathbf{F}_{\pm}^{(n)}(\mathbf{X}_i) = \mathbf{u}_{\kappa(i)}, \ i = 1, \ldots, n, \tag{4.4}$$

where $\kappa : \{1, 2, \ldots, n\} \to \{1, 2, \ldots, n\}$ denotes the coupling of the observed data $(\mathbf{X}_1, \ldots, \mathbf{X}_n)$ to the uniformly distributed points $\mathbf{U}_1, \ldots, \mathbf{U}_n$ on the sphere $\mathbb{S}_d$. Glivenko-Cantelli theorems were established without any moment conditions

Stellenbosch University https://scholar.sun.ac.za

for i.i.d. samples in Hallin (2017): as $n \to \infty$

$$\max_{1\le i\le n} \left\| \mathbf{F}_{\pm}^{(n)}(\mathbf{X}_i) - \mathbf{F}_{\pm}(\mathbf{X}_i) \right\| \to_{a.s.} 0.$$

In order to extend this transport over all $\mathbf{x} \in \mathbb{R}^d$ or $\mathbf{u} \in \mathbb{S}_d$, an interpolation is needed. Here, we concentrate on the extension of the map from $\mathbb{S}_d$ to $\mathbb{R}^d$. Such interpolations should belong to the class of gradients $\mathbf{T}_n : \mathbb{S}_d \to \mathbb{R}^d$ of convex functions Ψ_n, so that the resulting contours should be nested in order to have the true nature of (continuous) quantile contours. This program was carried out in del Barrio et al. (2018), where a smooth extension of (4.4) is proposed, and a construction was provided of the convex function Ψ_n whose gradient equals the transport $\mathbf{T}_n$ pushing the uniform distribution on $\mathbb{S}_d$ to the constructed empirical distribution on $\mathbb{R}^d$. In the sequel we also briefly write $\mathbf{F}_n$ for $\mathbf{F}_{\pm}^{(n)}$.

In the next section, we first construct an alternative smooth extension $\mathbf{T}_{n,\xi}$ of (4.4), depending on a smoothing parameter ξ, which allows for an integral representation of the volumes of the corresponding quantile contours, and we construct the corresponding potential $\Psi_{n,\xi}$. Then we consider the estimation of risk measures respectively based on $\Psi_{n,\xi}$, $\mathbf{T}_{n,\xi}$ and the volumes of the constructed quantile contours. The potential then leads to generalizations of risk measures based on the (univariate) integrated quantile functions as discussed in detail in Gushchin and Borzykh (2018). On the other hand, use of $\mathbf{T}_{n,\xi} = \nabla\Psi_{n,\xi}$ leads to estimation of maximal correlation risk measures of the type $\mathbb{E}\left(\langle \mathbf{U}, \nabla\Psi(\mathbf{U})\rangle\right)$ as discussed in Ekeland et al. (2012).

We also discuss the computation of the volumes of the empirical contours that can serve as empirical univariate quantiles and risk measures. Finally, we discuss some applications of the transport construction to risk measurement for multivariate regularly varying distributions. While the theoretical (asymptotic) characteristics of such statistics are outside the scope of this paper, we provide simulation results and some applications from insurance and finance, to show the usefulness of the proposed methods.

Stellenbosch University https://scholar.sun.ac.za

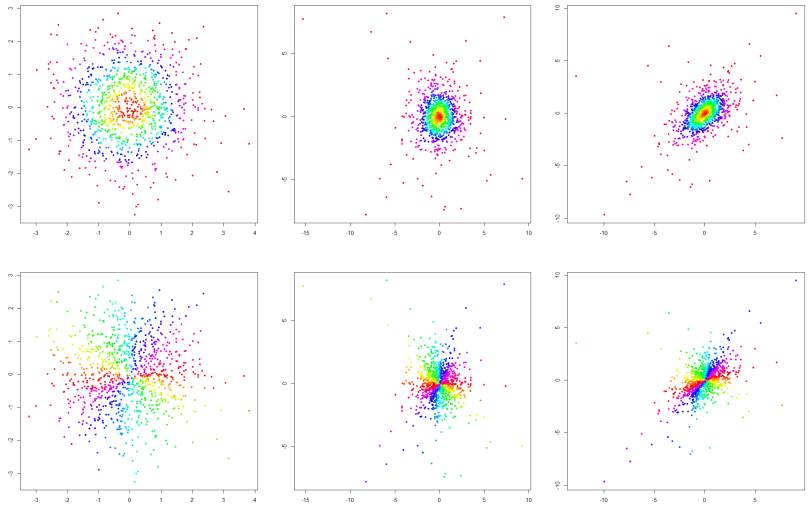

Figure 4.1: The coupling of bivariate normal (left), t_3 (center) and elliptical hyperbolic distribution samples with $\gamma = 1/3$ (right) where the colours illustrate quantile contours (top) and directions (bottom). Each sample is of size 1000.

4.2 Cyclically monotone interpolation

del Barrio et al. (2018) introduce a cyclically monotone map $\mathbf{T}_n : \mathbb{S}_d \to \mathbb{R}^d$ as follows: for $\mathbf{u} = \mathbf{u}_1, \ldots, \mathbf{u}_n$

$$\begin{aligned}\mathbf{T}_n(\mathbf{u}) &= \nabla \max_{i=1,2,\ldots,n} \psi_i(\mathbf{u}) \\ &= \mathbf{x}_i \quad \text{such that} \quad \psi_i(\mathbf{u}) > \psi_j(\mathbf{u}) \quad \text{for all} \quad i \neq j \in \{1, 2, \ldots, n\}\end{aligned}$$

where $\psi_i(\mathbf{u}) = \langle \mathbf{u}, \mathbf{x}_i \rangle - \lambda_i$, and $\lambda_1 \ldots \lambda_n \in \mathbb{R}$ are constants satisfying

$$\psi_i(\mathbf{u}_i) - \psi_j(\mathbf{u}_i) > 0 \quad \text{for all} \quad i \neq j \in \{1, 2, \ldots, n\}.$$

This is equivalent to satisfying $\langle \mathbf{u}_i, \mathbf{x}_i - \mathbf{x}_j \rangle > \lambda_i - \lambda_j$ for all $i \neq j \in \{1, 2, \ldots, n\}$, or

$$\langle \mathbf{u}_i, \mathbf{x}_i \rangle - \lambda_i > \max_{j=1,\ldots,n; j \neq i} (\langle \mathbf{u}_i, \mathbf{x}_j \rangle - \lambda_j). \tag{4.5}$$

Stellenbosch University https://scholar.sun.ac.za

It then follows that the constants $\lambda_1, \lambda_2, \ldots, \lambda_n$ are the solutions to the linear program maximizing δ such that

$$\begin{bmatrix} 1 & -1 & 0 & \ldots & 0 & 1 \\ 1 & 0 & -1 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 1 & 0 & 0 & \ldots & -1 & 1 \\ -1 & 1 & 0 & \ldots & 0 & 1 \\ 0 & 1 & -1 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 1 & 0 & \ldots & -1 & 1 \\ -1 & 0 & 1 & \ldots & 0 & 1 \\ 0 & -1 & 1 & \ldots & 0 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ -1 & 0 & 0 & \ldots & 1 & 1 \\ 0 & -1 & 0 & \ldots & 1 & 1 \\ 0 & 0 & -1 & \ldots & 1 & 1 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \ldots & 1 & 1 \end{bmatrix} \begin{bmatrix} \lambda_1 \\ \lambda_2 \\ \vdots \\ \lambda_n \\ \delta \end{bmatrix} \leq \begin{bmatrix} \langle \mathbf{u}_1, \mathbf{x}_1 - \mathbf{x}_2 \rangle \\ \langle \mathbf{u}_1, \mathbf{x}_1 - \mathbf{x}_3 \rangle \\ \vdots \\ \langle \mathbf{u}_1, \mathbf{x}_1 - \mathbf{x}_n \rangle \\ \langle \mathbf{u}_2, \mathbf{x}_2 - \mathbf{x}_1 \rangle \\ \langle \mathbf{u}_2, \mathbf{x}_2 - \mathbf{x}_3 \rangle \\ \vdots \\ \langle \mathbf{u}_2, \mathbf{x}_2 - \mathbf{x}_n \rangle \\ \langle \mathbf{u}_3, \mathbf{x}_3 - \mathbf{x}_1 \rangle \\ \langle \mathbf{u}_3, \mathbf{x}_3 - \mathbf{x}_2 \rangle \\ \vdots \\ \langle \mathbf{u}_n, \mathbf{x}_n - \mathbf{x}_1 \rangle \\ \langle \mathbf{u}_n, \mathbf{x}_n - \mathbf{x}_2 \rangle \\ \langle \mathbf{u}_n, \mathbf{x}_n - \mathbf{x}_3 \rangle \\ \vdots \\ \langle \mathbf{u}_n, \mathbf{x}_n - \mathbf{x}_{n-1} \rangle \end{bmatrix}$$

Note that δ is then defined as

$$\delta = \min_{i=1,\ldots,n} \left\{ \psi_i(\mathbf{u}_i) - \max_{j \neq i} \psi_j(\mathbf{u}_i) \right\} > 0,$$

the minimum difference between $\psi_i(\mathbf{u}_i)$ and $\psi_j(\mathbf{u}_i)$ for all $i \neq j$.

Considering the convex map Ψ_n defined as

$$\mathbf{u} \mapsto \Psi_n(\mathbf{u}) := \max_{i=1,\ldots,n} (\langle \mathbf{u}, \mathbf{x}_i \rangle - \lambda_i),$$

and the open convex sets $C_i = \{\mathbf{u} \in \mathbb{S}_d \,|\, (\langle \mathbf{u}, \mathbf{x}_i \rangle - \lambda_i) > \max_{j \neq i} (\langle \mathbf{u}, \mathbf{x}_j \rangle - \lambda_j)\}$ such that Ψ_n is differentiable in C_i, with $\nabla \Psi_n(\mathbf{u}) = \mathbf{x}_i$ for $\mathbf{u} \in C_i$, the map $\mathbf{T}_n$ can be extended to $\cup_{i=1}^n C_i$ by setting $\mathbf{T}_n(\mathbf{u}) = \nabla \Psi_n(\mathbf{u}),\ \mathbf{u} \in \cup_{i=1}^n C_i$.

del Barrio et al. (2018) constructed a continuous version of $\mathbf{T}_n$ by placing a Moreau envelope on Ψ_n: for some $\epsilon > 0$

$$\tilde{\mathbf{T}}_{n,\epsilon}(\mathbf{u}) = \nabla \Psi_{n,\epsilon}(\mathbf{u}) \text{ where } \Psi_{n,\epsilon}(\mathbf{u}) = \inf_{\mathbf{v} \in \mathbb{S}_d} \left\{ \Psi_n(\mathbf{v}) + \frac{1}{2\epsilon} \|\mathbf{u} - \mathbf{v}\|^2 \right\}.$$

Stellenbosch University https://scholar.sun.ac.za

This is equivalent to

$$\tilde{\mathbf{T}}_{n,\epsilon}(\mathbf{u}) = \sum_{i=1}^{n} w_{i,\epsilon}(\mathbf{u})\mathbf{x}_i$$

where $w_{i,\epsilon},\ i = 1, \ldots, n$ are a solution to the maximization problem

$$\max_{w_1,\ldots,w_n} \left(\sum_{i=1}^{n} w_i \psi_i(\mathbf{u}) - \frac{\epsilon}{2} \left\| \sum_{i=1}^{n} w_i \mathbf{x}_i \right\|^2 \right) \text{ such that } 0 \leq w_i \leq 1 \text{ and } \sum_{i=1}^{n} w_i = 1,$$

which can be solved by using a gradient descent algorithm. It was then shown that for sufficiently small $\epsilon > 0$, $\tilde{\mathbf{T}}_{n,\epsilon}$ provides a continuous, cyclically monotone interpolation of $(\mathbf{u}_i, \mathbf{x}_i),\ i = 1, \ldots, n$.

Here we propose an alternative continuous transformation

$$\begin{aligned} \mathbf{T}_{n,\xi}(\mathbf{u}) &= \frac{1}{\xi} \nabla \Big(\log \sum_{i=1}^{n} e^{\xi \psi_i(\mathbf{u})} \Big) \\ &= \frac{\sum_{i=1}^{n} e^{\xi\, \psi_i(\mathbf{u})}\, \mathbf{x}_i}{\sum_{i=1}^{n} e^{\xi\, \psi_i(\mathbf{u})}} \\ &= \sum_{i=1}^{n} w_{i,\xi}(\mathbf{u})\, \mathbf{x}_i, \end{aligned} \tag{4.6}$$

where $\xi > 0$ is chosen as large as possible and $w_{i,\xi}(\mathbf{u}) = \frac{e^{\xi\, \psi_i(\mathbf{u})}}{\sum_{m=1}^{n} e^{\xi\, \psi_m(\mathbf{u})}}$.
The potential function

$$\Psi_{n,\xi}(\mathbf{u}) = \frac{1}{\xi} \log \sum_{i=1}^{n} e^{\xi \psi_i(\mathbf{u})} \tag{4.7}$$

is convex as it is the logarithm of a sum of exponentials of convex functions, see Example 3.14, p. 87 in Boyd and Vandenberghe (2004). Hence $\mathbf{T}_{n,\xi}$ is a cyclically monotone map and the quantile contours corresponding to $\mathbf{T}_{n,\xi}(\mathbf{u})$ are nested. Note also that taking $\xi \to \infty$ brings this smoothed transport potential to the 'non-smoothed' potential $\mathbf{T}_n$, while as $\xi \to 0$, $\mathbf{T}_{n,\xi}$ approaches the constant function $\mathbf{u} \mapsto \frac{1}{n} \sum_{i=1}^{n} \mathbf{x}_i$, being the ultimate smooth version.

For convenience, we collect some key properties of $\Psi_{n,\xi}$ in the next result. Observe that in Proposition 4.1, *2.* follows from the last comment, while *1.* is

Stellenbosch University https://scholar.sun.ac.za

an elementary bound.

Proposition 4.1. $\Psi_{n,\xi}$ from (4.7) with $\psi_1, \ldots, \psi_n$ defined in (4.5) has the following properties:

1. $0 \leq \Psi_{n,\xi}(\mathbf{u}) - \Psi_n(\mathbf{u}) \leq \frac{\log n}{\xi}$

2. $\Psi_{n,\xi}$ is convex and differentiable.

Under mild regularity assumptions on the distribution of the $\mathbf{X}_i$, the center-outward distribution function, $\mathbf{F}_{\pm}$, and its inverse, the quantile function $\mathbf{T}_{\pm}$, are continuous ($\mathbf{T}_{\pm}$ is continuous over the unit ball except, possibly, at the origin, see Hallin (2017)). A sufficient condition ensuring this is that the law of the $\mathbf{X}_i$ belongs to the class $\mathcal{P}_d$ of probabilities with a non-vanishing density, f, over $\mathbb{R}^d$ such that for every $D > 0$ there exist constants $0 < \lambda_{f,D} < \Lambda_{f,D}$ such that

$$\lambda_{f,D} < f(\mathbf{x}) < \Lambda_{f,D}$$

for all $\|\mathbf{x}\| \leq D$. Under this assumption and a proper choice of the smoothing parameter ξ we can guarantee consistency of the smoothed empirical quantiles to their population counterparts. This is the content of our main result, a kind of Glivenko-Cantelli theorem for center-outward quantiles. We note that we may have to dismiss the outer parts of the unit ball to ensure uniform convergence to the population quantiles. This behavior already shows up for classical quantiles in dimension one, where uniform convergence may require to consider intervals of type $[a, b] \subset (0, 1)$.

Theorem 2.2. If the law of the $\mathbf{X}_i$ belongs to $\mathcal{P}_d$ and ξ_n satisfies

$$\lim_{n\to\infty} \frac{\xi_n}{\log n} = \infty$$

then, for every compact $K \subset B(\mathbf{0}, 1) - \{\mathbf{0}\}$,

$$\sup_{\mathbf{u}\in K} \|\mathbf{T}_{n,\xi_n}(\mathbf{u}) - \mathbf{T}(\mathbf{u})\| \to 0 \quad \text{a.s.}.$$

In particular, on a probability one set, $\mathbf{T}_{n,\xi_n}(\mathbf{u}) \to \mathbf{T}(\mathbf{u})$ for every $\mathbf{u} \in B(\mathbf{0}, 1) - \{\mathbf{0}\}$.

Proof. We write $\mathbf{P}$ for the law of $\mathbf{X}_i$ and $\mathbf{P}_n$ for the empirical measure on the $\mathbf{X}_i$. Then $\mathbf{P}_n \to_w \mathbf{P}$ on a probability one set. We denote as Ψ an optimal transportation potential from $\mathbf{U}_d$ to P, that is, a convex, lower semi-continuous

Stellenbosch University https://scholar.sun.ac.za

function differentiable at every point of the open unit ball (except, possibly, at the origin) and such that $\nabla\Psi = \mathbf{T}$. We note that existence of the convex potential Ψ follows from general duality theory and also that Ψ is uniquely defined up to the addition of a constant, see del Barrio and Loubes (2019) for details. Since we assume that $\mathbf{P} \in \mathcal{P}_d$ we also have continuity of $\mathbf{T}$ and, as a consequence, Ψ is differentiable at every point in the open unit ball (except, possibly, at the origin).

We apply now Theorem 2.8 in del Barrio and Loubes (2019) to conclude that there exist some constants, a_n, such that $\Psi_n(\mathbf{u}) - a_n \to \Psi(\mathbf{u})$ for every $\mathbf{u}$ in the (punctured) open unit ball. We note that, while the statement of the cited result assumes convergence in transportation cost metric rather than weak convergence, the proof depends only on the fact that, in that case, $\pi_n = (\nabla\Psi_n \times Id)\sharp\mathbf{P}_n$ (the probability induced from $\mathbf{P}_n$ through the map $(\nabla\Psi_n \times Id)$) converges weakly to $\pi = (\nabla\Psi \times Id)\sharp\mathbf{P}$, which holds in the setup considered here, see del Barrio et al. (2018). The assumption (4.3) ensures now that also $\Psi_{n,\xi_n}(\mathbf{u}) - a_n \to \Psi(\mathbf{u})$ for every $\mathbf{u}$ in the punctured open unit ball. Hence, we can apply Theorem 25.7 in Rockaffelar (1996) to conclude that $\nabla(\Psi_{n,\xi_n}(\mathbf{u}) - a_n) = \mathbf{T}_{n,\xi_n}(\mathbf{u}) \to \mathbf{T}(\mathbf{u}) = \nabla\Psi_{n,\xi_n}(\mathbf{u})$ for every $\mathbf{u} \in B(0,1) - \{\mathbf{0}\}$, with uniform convergence in compact subsets of the punctured unit ball. This completes the proof. □

In the sequel we illustrate the proposed concepts using the multivariate normal, t and elliptical hyperbolic distribution defined by $(\gamma D)^{-1/2}\mathbf{Y}$ with D a $\chi^2_{1/\gamma}$ distributed random variable independent from the normal vector $\mathbf{Y} \sim \mathcal{N}_d(\mathbf{0}, \Sigma)$, with Σ a symmetric $d \times d$ matrix with diagonal elements equal to 1 and off-diagonal elements equal to 0.5.

Concerning the construction of uniform grids $(\mathbf{u}_1, \ldots, \mathbf{u}_n)$, in case $d = 2$ we use a polar coordinate construction

$$\mathbf{u}_i = \frac{i}{n+1}\begin{pmatrix}\cos\phi_i \\ \sin\phi_i\end{pmatrix},$$

with ϕ_i independent uniformly distributed on $(0, 2\pi)$, $i = 1, \ldots, n$.
In case $d \geq 3$ we use an empirical version of (4.2), i.e.

$$\frac{\mathbf{z}_i}{\|\mathbf{z}_i\|}\hat{F}_{R,n}(\|\mathbf{z}_i\|),\ i = 1, \ldots, n,$$

where $\mathbf{z}_i = (z_{1,i}, \ldots, z_{d,i})$ are independent d-dimensional normal vectors with independent standard normal components and $\hat{F}_{R,n}$ denotes the empirical distribution function based on the $\|\mathbf{z}_i\|$. We then add an averaging procedure

Stellenbosch University https://scholar.sun.ac.za

over M random grids in order to reduce the variability of the results, as discussed in the simulation and applications sections where we use $M = 10$.

In Figures 4.1 and 4.2 we illustrate the proposed coupling for simulated bivariate samples of size 1000 linking radii and quantile contours in the circle with the corresponding data points in the generated data sets. Note that in case of heavy tailed distributions (t and elliptical hyperbolic) the top contours with the most outlying observations exhibit an irregular star shaped form. The p^{th} quantile contour is defined as $C_{n,\xi,p} := \{\mathbf{T}_{n,\xi}(\mathbf{u}) \,|\, \|\mathbf{u}\| = p\}$, $p \in (0,1)$. In Figure 4.3 we plot the potential surfaces $\mathbf{u} \mapsto \Psi_{n,\xi}(\mathbf{u})$ for the samples used in Figures 4.1 and 4.2.

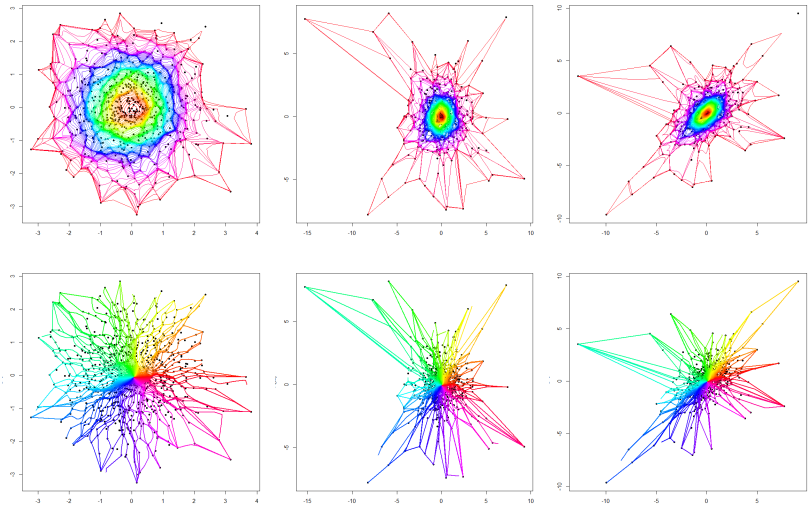

Figure 4.2: The sample quantile contours (top) and directions (bottom) based on $\mathbf{T}_{n,\xi}(\mathbf{u})$ for bivariate normal (left), t_3 (center) and elliptical hyperbolic distribution samples with $\gamma = 1/3$ (right). Each sample is of size 1000.

Stellenbosch University https://scholar.sun.ac.za

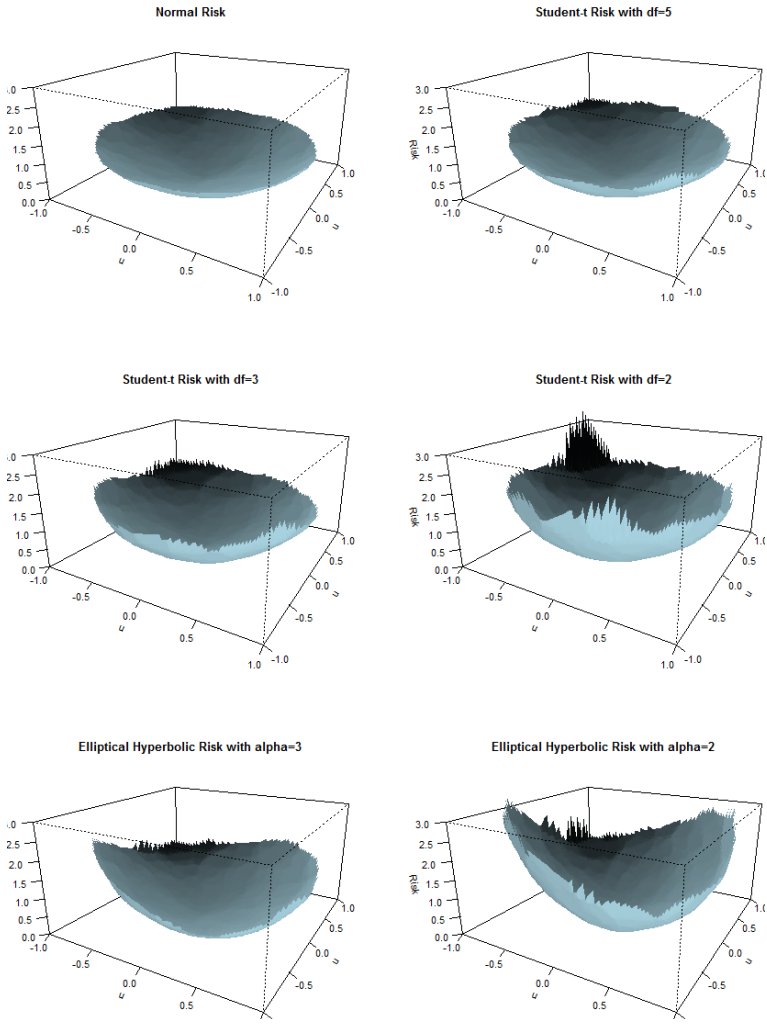

Figure 4.3: Potential surfaces of bivariate normal (top left), t_5 (top right), t_3 (center left) and t_2 (center right) samples with $\Sigma = \mathbf{I}_2$ and $\boldsymbol{\mu} = \mathbf{0}$, bivariate elliptical hyperbolic distribution samples with $\alpha = 1/\gamma = 3$ (bottom left) and $\alpha = 1/\gamma = 2$ (bottom right). Each sample is of size 1000.

Stellenbosch University https://scholar.sun.ac.za

4.3 Risk measurement based on Ψ_n and $\mathbf{T}_n$

In the univariate case Gushchin and Borzykh (2018) discussed integrated quantile and distribution functions which occur in risk measurement. Here we generalize the regular coherent risk measure $\int_0^1 u\,Q(u)\,du$ (see Kusuoka (2001)). To this end we consider the risk surface

$$\mathbf{u} \mapsto \mathbf{u} \cdot \nabla \Psi_{n,\xi}(\mathbf{u}), \quad \text{for } \mathbf{u} \in \mathbb{S}_d$$

which will be used as a basis for multivariate risk measurement. In Figure 4.4 this convex risk surface is plotted for bivariate normal, Student-t and elliptical hyperbolic distribution samples of size 1000. Note the difference in curvature between these cases.

Rüschendorf (2006) and Ekeland et al. (2012) define maximal correlation risk measures with respect to a baseline distribution μ as

$$\rho(\mathbf{X}) = \sup\{\mathbb{E}(\mathbf{X} \cdot \tilde{\mathbf{U}}) : \tilde{\mathbf{U}} \sim \mu\}.$$

By choosing the baseline distribution as the uniform distribution on the sphere $\mu = U$, it follows from Appendix B in Ekeland et al. (2012) that the maximal correlation risk measure corresponding to Ψ_n is given by

$$\rho_n(\mathbf{X}) = \mathbb{E}\left(\mathbf{U} \cdot \nabla \Psi_n(\mathbf{U})\right),$$

which can be estimated by

$$\hat{\rho}_n(\mathbf{X}) := \frac{1}{n} \sum_{i=1}^{n} \langle \mathbf{u}_{\kappa(i)}, \mathbf{x}_i \rangle.$$

In Table 4.1 the values of $\hat{\rho}_n$ are given for simulated samples of standard multivariate normal and t distributions for different dimensions. In each instance the average $\hat{\rho}_n$ is taken. Note the increase when increasing tail heaviness and dimensions.

The effect of outlying observations can be amplified by restricting this measure to the data which are lying outside a high quantile contour $R_{1-p} := \{\mathbf{T}_{n,\xi}(\mathbf{u}) : \|\mathbf{u}\| > 1 - p\}$ with $p \in (0, 1)$ small, for instance $p = 0.05$. This then leads to the empirical risk measure

$$\rho_{n,p}(\mathbf{X}) = \mathbb{E}\left([\mathbf{U} \cdot \nabla \Psi_n(\mathbf{U})] | \|\mathbf{U}\| > 1 - p\right),$$

Stellenbosch University https://scholar.sun.ac.za

Table 4.1: Summary of $\hat{\rho}_n$ values for samples of different dimensions, each of size 1000

Dimension	Normal	t_5	t_3
2	0.80	1.00	1.21
3	0.97	1.23	1.48
4	1.08	1.37	1.66
5	1.16	1.47	1.77

which can be estimated by

$$\hat{\rho}_{n,p}(\mathbf{X}) := \frac{1}{\sum_{i=1}^{n} 1_{[\|\mathbf{u}_i\|>1-p]}} \sum_{i=1}^{n} \langle \mathbf{u}_{\kappa(i)}, \mathbf{x}_i \rangle \, 1_{[\|\mathbf{u}_{\kappa(i)}\|>1-p]} \text{ as } \xi \to \infty.$$

Note that $\hat{\rho}_{n,p}$ is again a maximum correlation risk measure based on the convex potential $(\Psi_n(\mathbf{u}) - c, 0)$ where c is the value of $\Psi_n(\mathbf{u})$ on the $1-p$ quantile contour.

Table 4.2: Summary of $\hat{\rho}_{n,0.05}$ values for samples of different dimensions, each of size 1000

Dimension	Normal	t_5	t_3
2	2.61	4.17	6.30
3	2.69	4.37	6.48
4	2.73	4.30	6.35
5	2.78	4.35	6.17

Stellenbosch University https://scholar.sun.ac.za

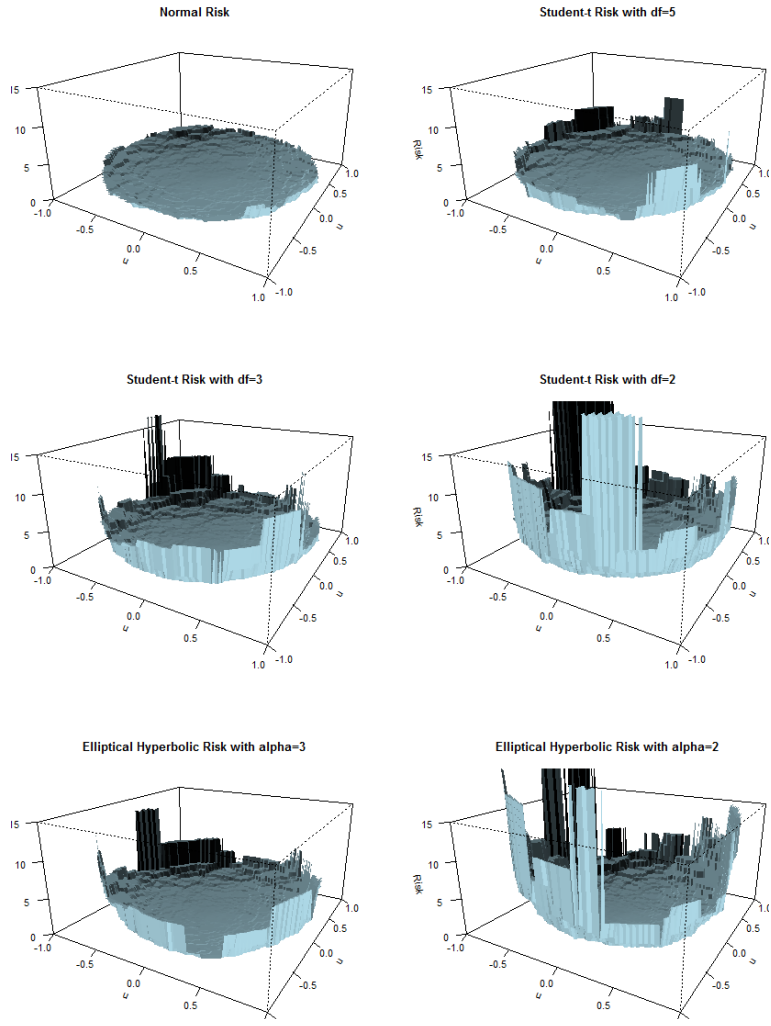

Figure 4.4: Risk surfaces of bivariate normal (top left), t_5 (top right), t_3 (center left) and t_2 (center right) samples with $\Sigma = \mathbf{I}_2$ and $\boldsymbol{\mu} = \mathbf{0}$, bivariate elliptical hyperbolic distribution samples with $\alpha = 1/\gamma = 3$ (bottom left) and $\alpha = 1/\gamma = 2$ (bottom right). Each sample is of size 1000.

Stellenbosch University https://scholar.sun.ac.za

4.4 On volumes of quantile regions

In this section we discuss statistical risk measurement based on the *quantile volumes* V_p of the quantile regions $R_{n,\xi,p} := \{\mathbf{T}_{n,\xi}(\mathbf{u}) \,|\, \|\mathbf{u}\| \leq p\}$, $p \in (0,1)$. Using polar coordinates we obtain

$$\begin{aligned}\hat{V}_n(p) &= \int_{R_p} dx_1\, dx_2 \ldots dx_d \\ &= \int_{\mathbf{u}:\|\mathbf{u}\|\leq p} J(\mathbf{u})\, du_1\, du_2 \ldots du_d \\ &= \int_0^p \int_0^{2\pi} \int_0^{\pi} \ldots \int_0^{\pi} J_{n,\xi}(r\,\mathbf{v}(\boldsymbol{\phi}))\, r^{d-1} \prod_{j=2}^{d-1} \sin^{d-j}\phi_j \, d\phi_1 \ldots d\phi_{d-1}\, dr\end{aligned}$$

where

$$\mathbf{v}(\boldsymbol{\phi}) = \begin{bmatrix} cos\phi_1 \\ sin\phi_1 cos\phi_2 \\ sin\phi_1 sin\phi_2 \cos\phi_3 \\ \vdots \\ sin\phi_1 \ldots sin\phi_{d-2} cos\phi_{d-1} \\ sin\phi_1 \ldots sin\phi_{d-2} sin\phi_{d-1} \end{bmatrix}$$

and

$$\begin{aligned}J_{n,\xi}(\mathbf{u}) &= |\nabla \mathbf{T}_{n,\xi}(\mathbf{u})| \\ &= \xi^d \left| \sum_{i=1}^{n} w_{i,\xi}\,(\mathbf{u})\,(\mathbf{x}_i - \bar{\mathbf{x}})\,\mathbf{x}_i^T \right| \\ &= \xi^d \left| \left(\mathbf{X} - \mathbf{1}\,\bar{\mathbf{x}}^T\right)^T W_\xi(\mathbf{u})\, \left(\mathbf{X} - \mathbf{1}\,\bar{\mathbf{x}}^T\right) \right|\end{aligned}$$

where $w_{i,\xi}(\mathbf{u})$ is from (4.6), $W_\xi(\mathbf{u}) = \operatorname{diag}\{w_{1,\xi}(\mathbf{u}), w_{2,\xi}(\mathbf{u}), \ldots, w_{n,\xi}(\mathbf{u})\}$, $\mathbf{X} = [\mathbf{x}_1\, \mathbf{x}_2\, \ldots\, \mathbf{x}_n]^T$, $\bar{x}_j = \sum_{i=1}^n w_{i,\xi}(\mathbf{u})\, x_{i,j}$, and $\bar{\mathbf{x}} = [\bar{x}_1\, \bar{x}_2\, \ldots\, \bar{x}_d]^T$.

Note that $\left(X - \mathbf{1}\,\bar{\mathbf{x}}^T\right)^T W_\xi(\mathbf{u})\, \left(X - \mathbf{1}\,\bar{\mathbf{x}}^T\right)$ is a weighted covariance matrix.

Stellenbosch University https://scholar.sun.ac.za

For an elliptical distribution the theoretical expression for the volumes follows from (4.1), denoting the quantile function of the distribution of $\|\mathbf{Z}\|$ by Q_R:

$$
\begin{aligned}
V(p) &= \int \dots \int_{\mathbf{Q}(p\mathbb{S}_{d-1})} \prod_{i=1}^{d} dx_i \\
&= |\Sigma|^{1/2} \int \dots \int_{\{\mathbf{z}=\Sigma^{-1/2}\mathbf{Q}(\mathbf{u}):\|\mathbf{u}\|\le p\}} \prod_{i=1}^{d} dz_i \\
&= \frac{|\Sigma|^{1/2}}{d} \int_0^p \int_0^{2\pi} \int_0^{\pi} \dots \int_0^{\pi} \prod_{j=1}^{d-1} \sin^{d-1-j}\phi_j \; dr_z^d \, d\phi_1 \dots d\phi_{d-1} \\
&= \frac{|\Sigma|^{1/2}}{d} \int_0^p \int_0^{2\pi} \int_0^{\pi} \dots \int_0^{\pi} \prod_{j=1}^{d-1} \sin^{d-1-j}\phi_j \; dQ_R^d(r_u) \, d\phi_1 \dots d\phi_{d-1} \\
&= |\Sigma|^{1/2} \frac{\pi^{d/2}}{\Gamma(1+d/2)} Q_R^d(p). \qquad (4.8)
\end{aligned}
$$

Considering these volumes as generalized quantiles lead to a second set of risk measures. However in the case of heavy tailed distributions, given the star shaped shape of the top contours makes, generalized Value-at-Risk measures $V_n(p)$ with p close to 1, are less attractive. The generalized median $V(0.5)$ can serve as a reasonable risk indicator.

In case of a multivariate normal distribution the generalized empirical quantiles $V_n(\frac{i}{n+1})$ $(i = 1, \dots, n)$ based on (4.8) are expected to follow a $(\chi^2_d)^{d/2}$ distribution. Hence plotting the empirical quantiles $V_n(\frac{i}{n+1})$ against the corresponding $(\chi^2_d)^{d/2}$ quantiles yields a graphical goodness-of-fit test for multivariate normality, as illustrated in Figure 4.5.

In case of heavy tailed multivariate elliptical distributions such that $Q_R(1-p) = p^{-1/\alpha}\ell(1/p)$ for some $\alpha > 0$ and some slowly varying function ℓ (i.e. $\ell(tx)/\ell(t) \to 1$ as $t \to \infty$ for every $x > 0$), plotting $\log V_n(\frac{i}{n+1})$ against $-\log(1-\frac{i}{n+1})$ $(i = 1, \dots, n)$ should lead to linear patterns at high values $i/(n+1)$. However, due to the star shaped top contours, a concave bending appears at these values. This makes it less useful as a way to validate the presence of a heavy tail, or to estimate α. See Figure 4.6. Note also the important computational effort in constructing such QQ-plots.

Stellenbosch University https://scholar.sun.ac.za

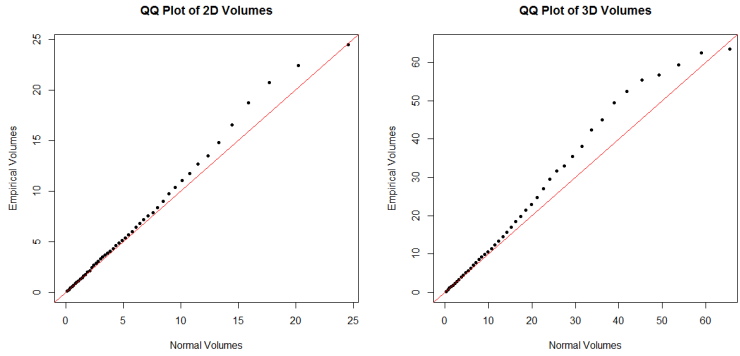

Figure 4.5: QQ-plots of the volumes $\hat{V}_n(p)$ from 2D (left) and 3D (right) standard normal sample against the standard normal volume quantiles. The red line represents the identity function, which gives an indication of how closely the volumes $\hat{V}_n(p)$ are to the actual standard normal volumes.

4.5 Analysis of Multivariate Regularly Varying Distributions

We consider multivariate regularly varying distributions for which there exist $\gamma > 0$ and a random vector $\mathbf{\Theta}$ such that for each $x > 0$ as $t \to \infty$

$$\frac{\mathbb{P}\left(\|\mathbf{X}\| > tx,\, \mathbf{X}/\|\mathbf{X}\| \in \cdot\right)}{\mathbb{P}\left(\|\mathbf{X}\| > t\right)} \to_v x^{-1/\gamma}\mathbb{P}(\mathbf{\Theta} \in \cdot). \tag{4.9}$$

The index of regular variation or extreme value index (EVI) $\gamma > 0$ itself is an important indicator of the heaviness of the multivariate tail, as the higher γ the higher the spread in the tails in the different directions.

We propose to evaluate the risk involved under (4.9) by applying Pareto QQ-plots based on the random variables $Y_i := \langle \mathbf{u}_{\kappa(i)}, \mathbf{X}_i \rangle$ for $i = 1, 2, \ldots, n$. The corresponding order statistics are denoted by $Y_{1,n} < Y_{2,n} < \cdots < Y_{n,n}$. The theoretical distribution $\tilde{Y}$ of the Y values (obtained for $n \to \infty$) can be conjectured to be of Pareto-type under (4.9), i.e.

$$\mathbb{P}(\tilde{Y} > y) = y^{-1/\gamma}\ell_F(y), \tag{4.10}$$

Stellenbosch University https://scholar.sun.ac.za

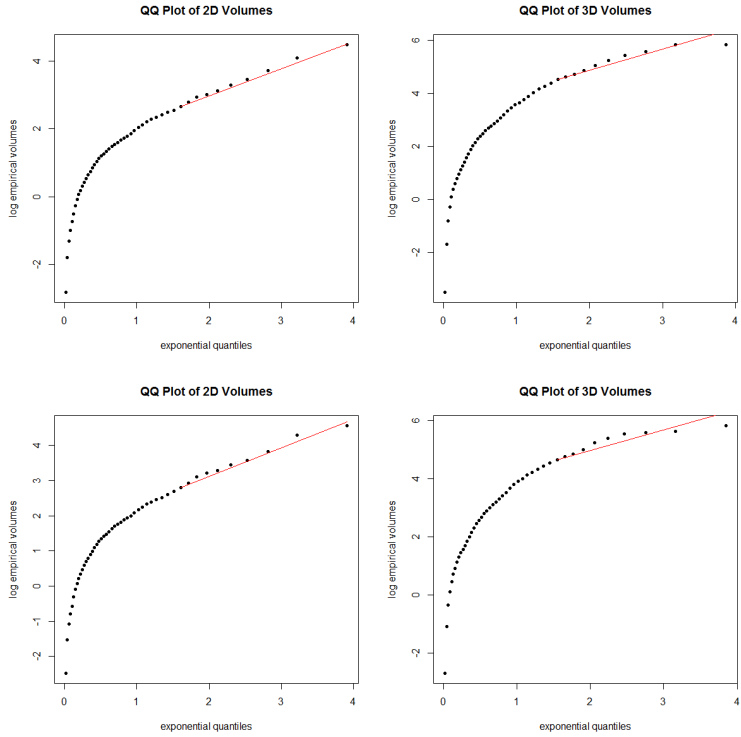

Figure 4.6: QQ-plots of the log-volumes $\log \hat{V}_n(p)$ from 2D (left) and 3D (right) elliptical hyperbolic samples (top) and Student-t samples (bottom) against the standard exponential quantiles. The red lines represent the gradient of the largest 10 QQ-points, which provide an indication of the extremity of the samples.

with ℓ_F some slowly varying function at infinity.
In the case of elliptical distributions, it follows from (4.2) that

$$\tilde{Y} = \langle \frac{\mathbf{U}}{\|\mathbf{U}\|}, \Sigma^{1/2} \frac{\mathbf{U}}{\|\mathbf{U}\|} \rangle \|\mathbf{U}\| \, Q_R(\|\mathbf{U}\|) + \langle \mathbf{U}, \boldsymbol{\mu} \rangle.$$

Here $Q_R(p) = (1-p)^{-\gamma} \ell_Q(1/(1-p))$ with ℓ_Q a slowly varying function and the factor $\frac{\langle \mathbf{U}, \Sigma^{1/2}\mathbf{U} \rangle}{\|\mathbf{U}\|}$ and the term $\langle \mathbf{U}, \boldsymbol{\mu} \rangle$ lead to further adjustments of the slowly varying function when $\|\mathbf{U}\| \to 1$, which depend on the direction of $\mathbf{U}$. Note that $\frac{\langle \mathbf{U}, \Sigma^{1/2}\mathbf{U} \rangle}{\|\mathbf{U}\|^2}$ is bounded below and above by the smallest and largest eigenvalue of $\Sigma^{1/2}$.

Stellenbosch University https://scholar.sun.ac.za

In the general multivariate regular variation case, (4.9) states that when restricting to radii above a certain threshold t, the transport of the exceedances $\mathbf{X}/t$ to $\mathbf{U}$ can also basically be split into a transport of the angular measure $\mathbf{\Theta}$ to the uniform distribution on $\mathcal{S}_{d-1}$ and the probability integral transform of the radii to the uniform (0,1) distribution, to be compared with (4.1). From this (4.10) can be conjectured.

Under (4.10), the Pareto QQ-plots

$$\left(\log(\frac{n+1}{j}), \log Y_{n-j+1,n}\right), \ j = 1, \ldots, n \tag{4.11}$$

are expected to show a linear pattern for large values of $\log Y$, say above the log-threshold point $\left(\log \frac{n+1}{k+1}, \log Y_{n-k,n}\right)$ for some $k \in \{2, \ldots, n-1\}$. Indeed, for $j = 1, \ldots, k+1$ the values $\log \tilde{Y}_{n-j+1,n}$ are approximately distributed as $\log Q_R(\|\mathbf{U}\|_{n-j+1,n})$ for $k, n \to \infty$ and $n/k \to \infty$ with expected value $\log Q_R(\frac{n-j+1}{n+1}) = \gamma \log(\frac{n+1}{j}) + \log \ell_Q(\frac{n+1}{j})$, and the slope in (4.11) over a high enough threshold $\log Y_{n-k,n}$ is expected to approximate γ. Assuming that $\ell_Q(x) = C(1 + o(1))$ as $x \to \infty$ with C the scale parameter of the Pareto-type distribution we obtain that the intercept of the fitted regression line indicates the value of $\log C$.

Formally, the slope can be measured by the average vertical increase in the Pareto QQ-plot above $\log Y_{n-k,n}$ using the approach proposed in Hill (1975):

$$H_{k,n} = \frac{1}{k} \sum_{j=1}^{k} \log \frac{Y_{n-j+1,n}}{Y_{n-k,n}} = \sum_{j=1}^{k} j \, \log \frac{Y_{n-j+1,n}}{Y_{n-j,n}}, \quad k = 1, \ldots, n.$$

Alternatively, bias reduction methods are available, such as the ridge regression method proposed by Buitendag et al. (2019):

$$\hat{\gamma}_k(\tau) = H_{k,n} - \frac{\bar{c}_k \, \sum_{j=1}^{k} (c_{j,k} - \bar{c}_k) \, j \, \log \frac{Y_{n-j+1,n}}{Y_{n-j,n}}}{\sum_{j=1}^{k} (c_{j,k} - \bar{c}_k)^2 + k \, \tau},$$

with $c_{j,k} = \left(\frac{j}{k+1}\right)^{-\rho}$ where $\rho < 0$. Note that the choice $\tau = 0$ of the ridge parameter τ leads to simple least squares regression $\hat{\gamma}_k^{\mathrm{LS}}$, see for instance Feuerverger and Hall (1999) and Beirlant et al. (1999).

In Figures 4.8 to 4.11 simulation results are reported showing that the proposed method based on the Y_i variables ($i = 1, \ldots, n$), show a promising behavior

Stellenbosch University https://scholar.sun.ac.za

as estimators of γ. For bivariate and trivariate elliptical hyperbolic and t distributions and sample size $n = 1000$ we report the bias and RMSE as a function of k for $H_{k,n}$, $\hat{\gamma}_k^{\mathrm{LS}} = \hat{\gamma}_k(0)$ and $\hat{\gamma}_k^{ridge} = \hat{\gamma}_k(\hat{\tau}_k)$, with $\hat{\tau}_k$ an estimator of the ridge parameter depending on an estimate of $b_{n,k}$ as proposed in Buitendag et al. (2019). These results compare favorably with the simulation results in Figure 1 in Kim and Lee (2017).

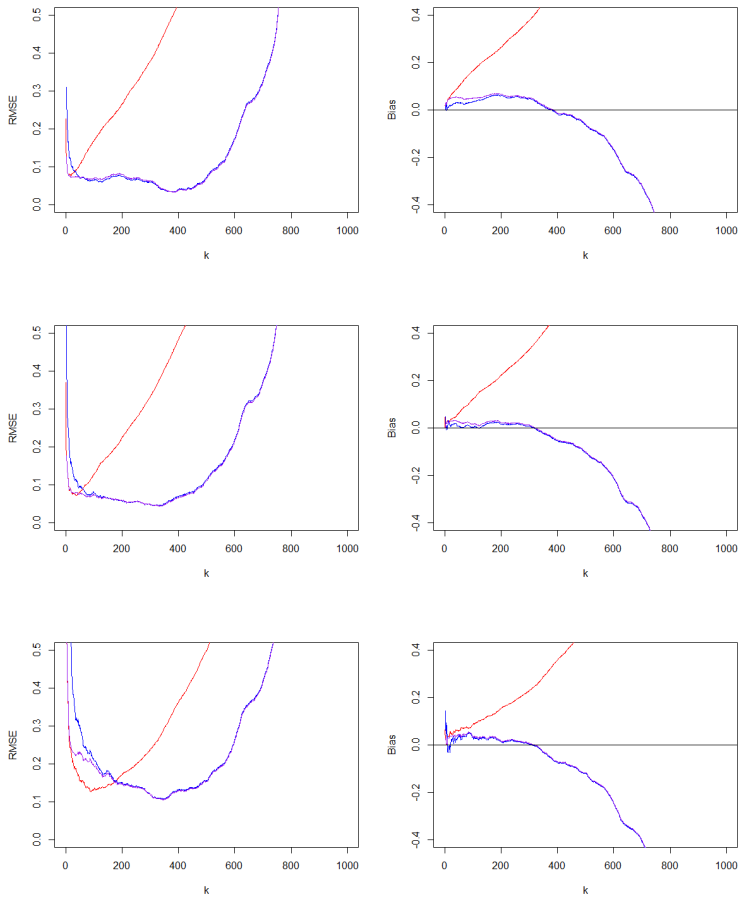

Figure 4.7: Bivariate elliptical hyperbolic distribution with an EVI of 1/5 (top), 1/3 (center) and 1 (bottom): RMSE (left) and bias (right) plots of the Hill (red), least-squares (blue) and ridge regression (purple) estimates for the EVI γ.

Stellenbosch University https://scholar.sun.ac.za

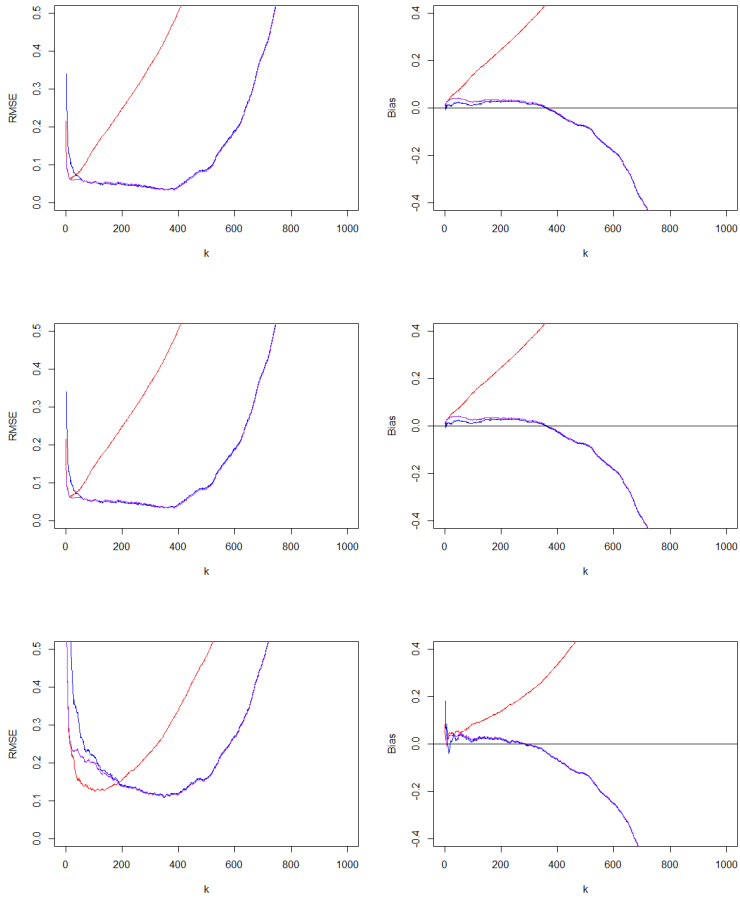

Figure 4.8: Bivariate Student-t distribution with an EVI of 1/5 (top), 1/3 (center) and 1 (bottom): RMSE (left) and bias (right) plots of the Hill (red), least-squares (blue) and ridge regression (purple) estimates for the EVI γ.

Stellenbosch University https://scholar.sun.ac.za

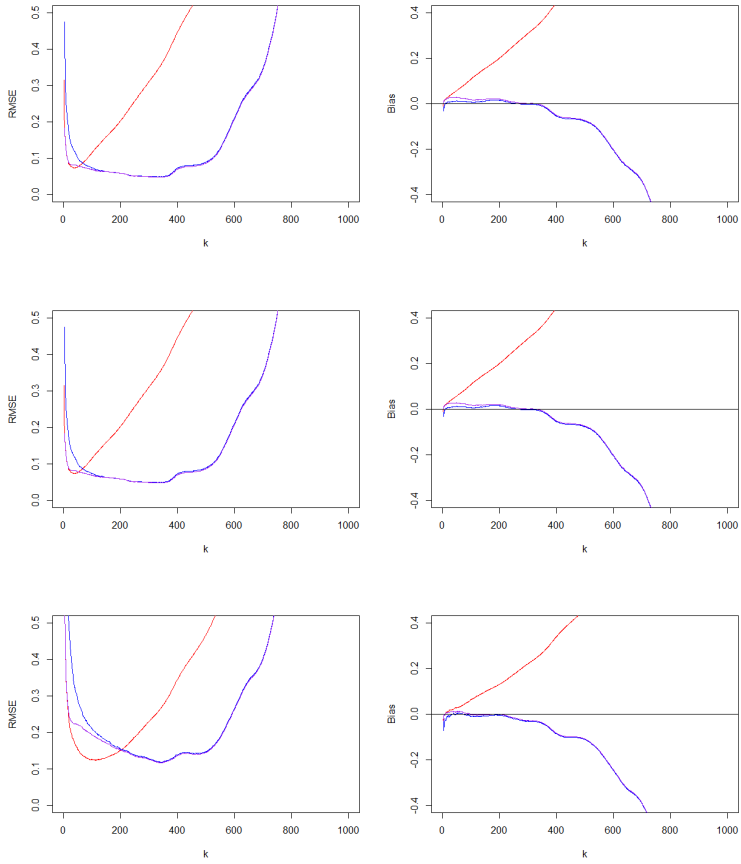

Figure 4.9: 3 dimensional elliptical hyperbolic distribution with an EVI of 1/5 (top), 1/3 (center) and 1 bottom: RMSE (left) and bias (right) plots of the Hill (red), least-squares (blue) and ridge regression (purple) estimates for the EVI γ.

Stellenbosch University https://scholar.sun.ac.za

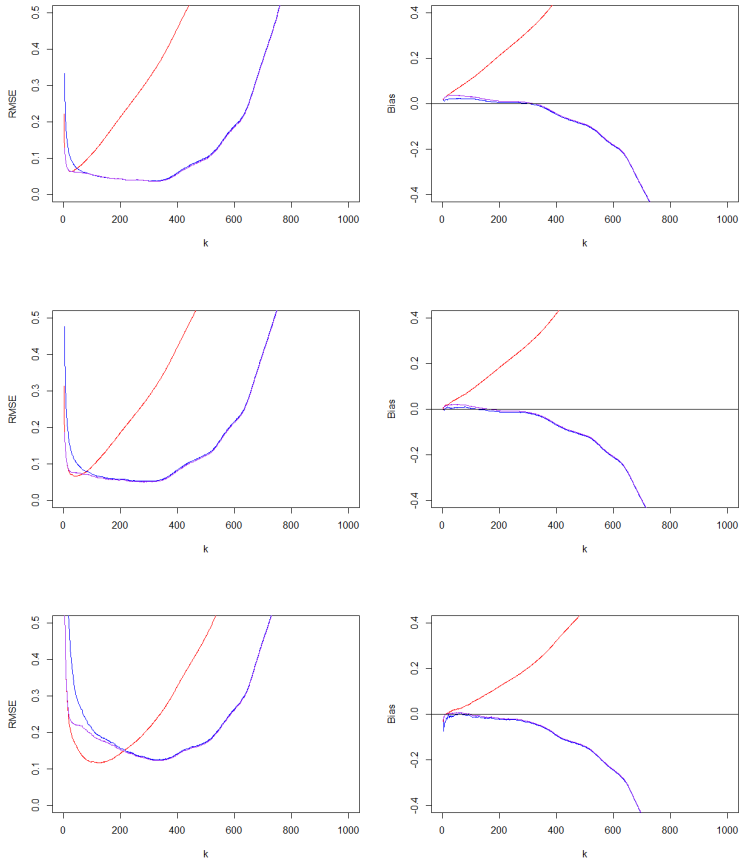

Figure 4.10: 3 dimensional Student-t distribution with an EVI of 1/5 (top), 1/3 (center) and 1 (bottom): RMSE (left) and bias (right) plots of the Hill (red), least-squares (blue) and ridge regression (purple) estimates for the EVI γ.

Stellenbosch University https://scholar.sun.ac.za

4.6 Case studies

4.6.1 Google Apple share log-returns

As a first case study we consider the daily log-returns of Google and Apple share prices for the period 01-01-2005 to 01-04-2019. We group three years' data at the end of the final month and denote these as 12-2007, 01-2008, up to 03-2019. For instance, the 12-2007 group consists of the data from 01-01-2005 to 31-12-2007, the 01-2008 group consists of the data from 01-02-2005 to 31-01-2008, and so forth.

Clearly when the 2008 financial crisis period has disappeared from the three year moving window, the risk surface is less high and the risk measures $\hat{\rho}_n$ reach a lower level. From the Pareto QQ-plots (4.11) and the corresponding EVI estimates we observe that the EVI does not really change over time, while the Pareto scale parameter decreases over time as can be seen in the Pareto QQ-plot from the decreasing height of the QQ-plot at higher exponential quantile levels.

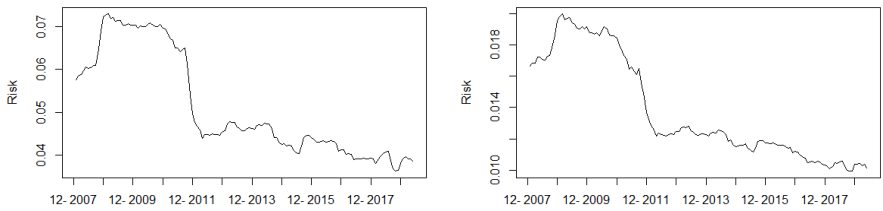

Figure 4.11: The multivariate risk measures $\hat{\rho}_n$ (left) and $\hat{\rho}_{n,0.1}$ (right) for the Google and Apple log-returns data for monthly periods from 2007 to 2019.

Stellenbosch University https://scholar.sun.ac.za

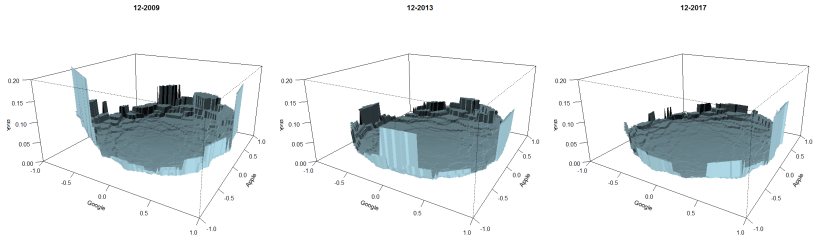

Figure 4.12: The bivariate risk surfaces for the Google and Apple log-returns data for the datasets 12-2009, 12-2013 and 12-2017.

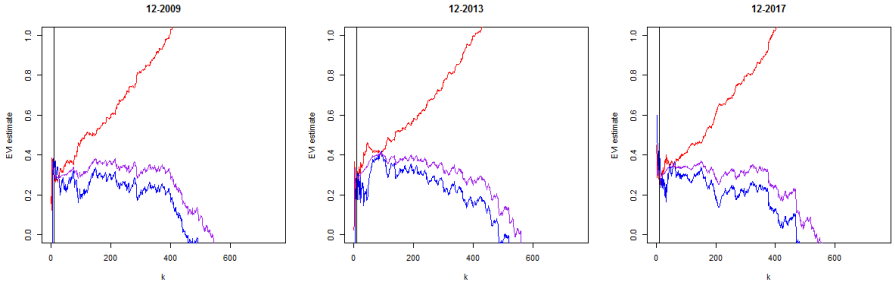

Figure 4.13: EVI estimates for the Google and Apple log-returns datasets 12-2009 (left), 12-2013 (center) and 12-2017 (right). The red, purple and blue lines represent the Hill, ridge regression and least squares EVI estimators respectively.

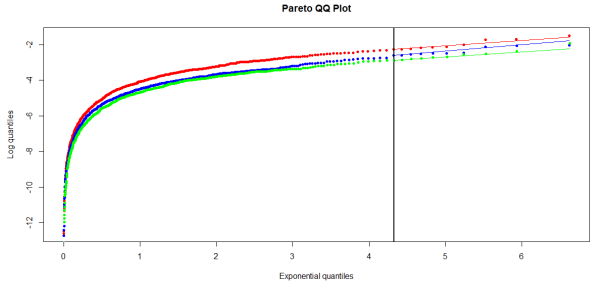

Figure 4.14: Pareto QQ-plots of $\langle \mathbf{u}_i, \mathbf{T}_n(\mathbf{u}_i) \rangle$ from the Google and Apple log-returns datasets 12-2009 (red), 12-2013 (blue) and 12-2017 (green)

Stellenbosch University https://scholar.sun.ac.za

4.6.2 Danish fire insurance

As a further case study we consider the log contents, building and profit amounts of the Danish fire insurance data for the period 1980 to 1990, discussed for instance in Embrechts et al. (1997) or Beirlant et al. (2004). We again group three years' data at the end of the final month and denote these as 12-1982, 01-1983, up to 12-1990. For instance, the 12-1982 group consists of the data from 01-01-1980 to 31-12-1982, the 01-1983 group consists of the data from 01-02-1980 to 31-01-1983, and so forth. Here the risk increases starting at the time period 1985-1988, while the EVI stays constant at 0.2 while the scale parameter increases.

Figure 4.15: The multivariate risk measures $\hat{\rho}_n$ (left) and $\hat{\rho}_{n,0.1}$ (right) for the Danish fire insurance data for monthly periods from 1980 to 1990.

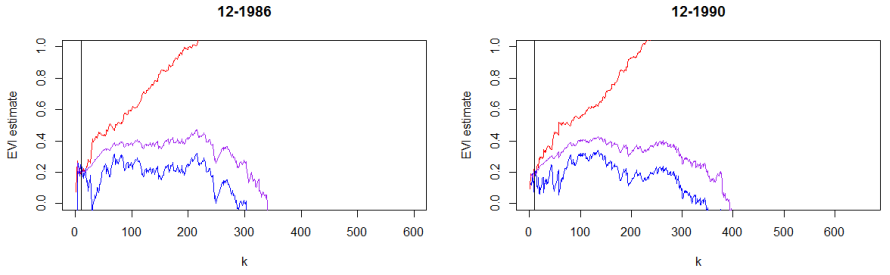

Figure 4.16: Potential-based EVI estimates for the Danish fire insurance datasets 12-1986 (left) and 12-1990 (right). The red, purple and blue lines represent the Hill, ridge regression and least squares EVI estimators respectively.

Stellenbosch University https://scholar.sun.ac.za

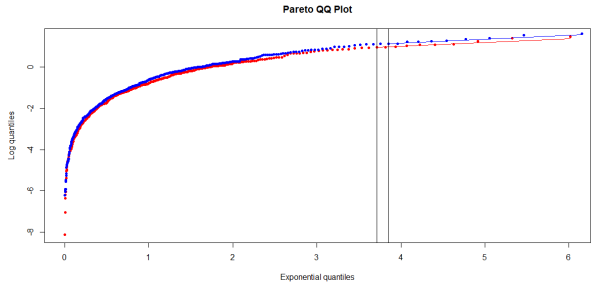

Figure 4.17: QQ-plots of the potential points $\langle \mathbf{u}_i, \mathbf{T}_n(\mathbf{u}_i) \rangle$ from the Danish fire insurance datasets 12-1986 (red) and 12-1990 (blue)

4.7 Conclusion

In this chapter we illustrate the use of recent transport methodology of center-outward distribution and quantile functions as introduced in Hallin (2017) and del Barrio et al. (2018) to multivariate risk measurement. Next to proposing a new interpolation method for the center-outward distribution function, we indicate several new multivariate risk measures, next to a novel estimator of the index of multivariate regular variation which exhibits a good behavior in simulations. Clearly, subsequent work has to focus on deriving consistency and the asymptotic distribution of the estimator of the index of multivariate regular variation, while other proposed ideas such as the principal components directions should be investigated in more depth.

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Appendix

Addendum A

Proof of Theorem 2.1. Note that

$$\hat{\gamma}_k^+ = \frac{1}{k}\sum_{j=1}^{k}\left\{1 - \bar{c}\frac{c_j - \bar{c}}{S_{cc} + \hat{\tau}_k^+}\right\} Z_j.$$

Using the consistency of $\hat{\beta}_k$ and $\hat{\rho}$ it follows that as $M \neq 0$ we have $\hat{\tau}_k^+ \to_p \gamma^2/M^2$ as $\sqrt{k}b_{n,k} \to M$. So

$$\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(c_j - \bar{c})\left(\frac{1}{S_{cc} + \hat{\tau}_k^+} - \frac{1}{S_{cc} + (\gamma/M)^2}\right) Z_j$$

$$=_d \frac{\hat{\tau}_k^+ - (\gamma/M)^2}{(S_{cc} + \hat{\tau}_k^+)(S_{cc} + (\gamma/M)^2)}\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(c_j - \bar{c})(\gamma + b_{n,k}c_j + \epsilon_j)$$

$$\to_p 0.$$

So the limit distribution is obtained from the limit distribution of

$$\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^{k}\left\{1 - \bar{c}\frac{c_j - \bar{c}}{S_{cc} + (\gamma/M)^2}\right\}(\gamma + b_{n,k}c_j + \epsilon_j) - \gamma\right).$$

In case $M = 0$ we have similarly that $(S_{cc} + \hat{\tau}_k^+)^{-1} \to_p 0$ so that we obtain the limit distribution of

$$\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^{k}(\gamma + b_{n,k}c_j + \epsilon_j) - \gamma\right). \qquad \square$$

Stellenbosch University https://scholar.sun.ac.za

Proof of Theorem 2.2. It follows from Beirlant et al. (2005) that the generalised log spacings $\{Y_j\}$ have the following asymptotic expansions

$$Y_j =_d \gamma\,(j+1)\log\left(1+\frac{1}{j}\right)+\left\{b_{n,k}\left(\frac{j}{k+1}\right)^{|\tilde{\rho}|}+\epsilon_j\right\}\{1+o_P(1)\},$$

where

$$\epsilon_j=\begin{cases}\frac{j+1}{\sqrt{k}}\left(\frac{k}{j}W^{(0)}\left(\frac{j}{k}\right)-\frac{k}{j+1}W^{(0)}\left(\frac{j+1}{k}\right)\right) & \text{if } \gamma>0,\\ \frac{(j+1)(1-\gamma)}{\sqrt{k}}\left(\left(\frac{k}{j}\right)^{1-\gamma}W^{(\gamma)}\left(\frac{j}{k}\right)-\left(\frac{k}{j+1}\right)^{1-\gamma}W^{(\gamma)}\left(\frac{j+1}{k}\right)\right) & \text{if } \gamma<0.\end{cases}$$

Furthermore, write $\{\lambda_j(\tau)\}$ as

$$\lambda_j(\tau)=\alpha(\tau)+\beta(\tau)\left(\frac{j}{k+1}\right)^{-\tilde{\rho}},$$

where $\alpha(\tau)=1+\frac{\bar{c}^2}{S_{cc}+\tau}$ and $\beta(\tau)=-\frac{\bar{c}}{S_{cc}+\tau}$. Also note that $\bar{c}\to\frac{1}{1-\tilde{\rho}}$ and $S_{cc}+\bar{c}^2\to\frac{1}{1-2\tilde{\rho}}$ as $k\to\infty$ so $\beta(\tau)=(1-\alpha(\tau))\,(1-\tilde{\rho})\,(1+o_p(1))$.

The proof of Theorem 2 follows then similarly as in the proof of Theorem 1, leading to the asymptotic distribution of

$$S_k:=\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^{k}\left\{1-\bar{c}\frac{c_j-\bar{c}}{S_{cc}+\tau_M}\right\}(\gamma+b_{n,k}c_j+\epsilon_j)-\gamma\right).$$

The expected value of the limit distribution of S_k is given by

$$b_{n,k}\bar{c}\frac{\tau_M}{S_{cc}+\tau_M}.$$

The asymptotic variance of S_k follows from a more tedious calculation, both in case $\gamma>0$ and $\gamma<0$.

The asymptotic variance when $\gamma>0$. The covariance of the error terms $\{\epsilon_j\}$ is given by

$$Cov(\epsilon_i,\epsilon_j)=\begin{cases}(1+\gamma^2)\left(1+\frac{1}{i}\right)-2\gamma\,(i+1)\log\left(1+\frac{1}{i}\right) & \text{if } i=j,\\ \gamma\,\frac{i+1}{j}\,\log\left(1+\frac{1}{i}\right) & \text{if } i<j.\end{cases}$$

Stellenbosch University https://scholar.sun.ac.za

It follows that the asymptotic variance of S_k is given by

$$\begin{aligned}
&\frac{1}{k^2} Var\left(\sum_{j=1}^{k} \lambda_j(\tau)\,\epsilon_j\right) \\
=&\frac{1}{k^2}\left(\sum_{j=1}^{k} \lambda_j^2(\tau)\,Var(\epsilon_j) + 2\sum_{i=1}^{k}\sum_{j=i+1}^{k} \lambda_i(\tau)\,\lambda_j(\tau)\,Cov(\epsilon_i,\epsilon_j)\right) \\
=&\frac{(1-\gamma)^2}{k}\left(1+\frac{\bar{c}^2 S_{cc}}{(S_{cc}+\tau_M)^2}\right) \\
&+\frac{2\gamma}{k}\left(1+\bar{c}^2 S_{cc}\left(\frac{\xi_1}{S_{cc}+\tau_M}\frac{\xi_2}{(S_{cc}+\tau_M)^2}\right)\right),
\end{aligned}$$

where $\xi_1 = \frac{-\tilde{\rho}\,\bar{c}}{S_{cc}}$ and $\xi_2 = \bar{c}$.

Indeed, as $k \to \infty$

$$\begin{aligned}
&\frac{1}{k}\sum_{j=1}^{k} \lambda_j^2(\tau_M)\,Var(\epsilon_j) \\
&= \frac{1}{k}\sum_{j=1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}\right)^2 \\
&\quad\times\left((1+\gamma^2)\left(1+\frac{1}{j}\right)-2\gamma\,(j+1)\log\left(1+\frac{1}{j}\right)\right) \\
&= \left\{(1-\gamma)^2\int_0^1 \left(\alpha(\tau_M)+\beta(\tau_M)\,u^{-\tilde{\rho}}\right)^2 du\right\}\{1+o(1)\} \\
&= (1-\gamma)^2\left\{\alpha^2(\tau_M)+\frac{2\alpha(\tau_M)\beta(\tau_M)}{1-\tilde{\rho}}+\frac{\beta^2(\tau_M)}{1-2\tilde{\rho}}\right\}\{1+o(1)\} \\
&= (1-\gamma)^2\left\{1+\frac{\bar{c}^2 S_{cc}}{(S_{cc}+\tau_M)^2}\right\}\{1+o(1)\},
\end{aligned}$$

Stellenbosch University https://scholar.sun.ac.za

and

$$
\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k}\sum_{j=i+1}^{k}\lambda_i(\tau_M)\,\lambda_j(\tau_M)\,Cov(\epsilon_i,\epsilon_j)\\
=&\frac{\gamma}{k^2}\sum_{i=1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{i}{k+1}\right)^{-\tilde{\rho}}\right)(i+1)\log\left(1+\frac{1}{i}\right)\\
&\times\sum_{j=i+1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}\right)\frac{k}{j}\\
=&\gamma\left\{\int_{\frac{1}{k}}^{1}\left(\alpha(\tau_M)+\beta(\tau_M)u^{-\tilde{\rho}}\right)(ku+1)\log\left(1+\frac{1}{ku}\right)\right.\\
&\left.\times\int_{u}^{1}\left(\alpha(\tau_M)+\beta(\tau_M)v^{-\tilde{\rho}}\right)\frac{dv}{v}\,du\right\}\{1+o(1)\}\\
=&\gamma\left\{\alpha^2(\tau_M)+\frac{\alpha(\tau_M)\beta(\tau_M)}{1-\tilde{\rho}}\left(1+\frac{1}{1-\tilde{\rho}}\right)+\frac{\beta^2(\tau_M)}{(1-\tilde{\rho})(1-2\tilde{\rho})}\right\}\{1+o(1)\}\\
=&\gamma\left\{1-\frac{\tilde{\rho}}{1-\tilde{\rho}}\frac{\bar{c}^2}{S_{cc}+\tau_M}+\frac{S_{cc}}{1-\tilde{\rho}}\frac{\bar{c}^2}{(S_{cc}+\tau_M)^2}\right\}\{1+o(1)\}
\end{aligned}
$$

Stellenbosch University https://scholar.sun.ac.za

The asymptotic variance when $\gamma < 0$. Here

$$Var(\epsilon_i) = \frac{1-\gamma}{1-2\gamma}\left\{(2-(1+\gamma)(1-2\gamma))\left(1+\frac{1}{i}\right)\right.$$
$$\left.+4(i+1)\left(1-\left(1+\frac{1}{i}\right)^{\gamma}\right)\right\}$$

and

$$Cov(\epsilon_i, \epsilon_j) = \frac{2(1-\gamma)}{1-2\gamma}(i+1)(j+1)\left(i^{-\gamma}-(i+1)^{-\gamma}\right)\left(j^{\gamma-1}-(j+1)^{\gamma-1}\right)$$

for $i < j$.

It follows that the asymptotic variance of S_k is given by

$$\frac{1}{k^2}Var\left(\sum_{j=1}^{k}\lambda_j(\tau_M)\,\epsilon_j\right)$$
$$=\frac{1}{k^2}\left(\sum_{j=1}^{k}\lambda_j^2(\tau_M)\,Var(\epsilon_j)+2\sum_{i=1}^{k}\sum_{j=i+1}^{k}\lambda_i(\tau_M)\,\lambda_j(\tau_M)\,Cov(\epsilon_i,\epsilon_j)\right)$$
$$=\frac{(1-\gamma)^2}{k}\left(1+\frac{\bar{c}^2 S_{cc}}{(S_{cc}+\tau_M)^2}\right)$$
$$+\frac{2\gamma}{k}\left(\frac{2(1-\gamma)}{1-2\gamma}+\bar{c}^2 S_{cc}\left(\frac{\xi_1(\gamma)}{S_{cc}+\tau_M}+\frac{\xi_2(\gamma)}{(S_{cc}+\tau_M)^2}\right)\right),$$

where $\xi_1(\gamma) = \frac{-2\tilde{\rho}(1-\gamma)}{S_{cc}(1-\gamma-\tilde{\rho})(1-2\gamma)}$ and $\xi_2(\gamma) = \frac{2(1-\gamma)^2}{(1-\gamma-\tilde{\rho})(1-2\gamma)}$.

Stellenbosch University https://scholar.sun.ac.za

Indeed, as $k \to \infty$

$$\frac{1}{k}\sum_{j=1}^{k}\lambda_j^2(\tau_M)\,Var(\epsilon_j)$$

$$=\frac{1-\gamma}{1-2\gamma}\frac{1}{k}\left\{(2-(1+\gamma)(1-2\gamma))\sum_{j=1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}\right)^2\left(1+\frac{1}{j}\right)\right.$$

$$\left.+4\sum_{j=1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}\right)^2(j+1)\left(1-\left(1+\frac{1}{j}\right)^{\gamma}\right)\right\}$$

$$=\frac{1-\gamma}{1-2\gamma}\left\{(2-(1+\gamma)(1-2\gamma))\int_{\frac{1}{k}}^{1}\left(\alpha(\tau_M)+\beta(\tau_M)u^{-\tilde{\rho}}\right)^2\left(1+\frac{1}{ku}\right)du\right.$$

$$\left.+4\int_{\frac{1}{k}}^{1}\left(\alpha(\tau_M)+\beta(\tau_M)u^{-\tilde{\rho}}\right)^2(ku+1)\left(1-\left(1+\frac{1}{ku}\right)^{\gamma}\right)du\right\}\{1+o(1)\}$$

$$=\frac{(1-\gamma)\,(2-(1+\gamma)(1-2\gamma)-4\gamma)}{1-2\gamma}\left\{\int_{\frac{1}{k}}^{1}\left(\alpha(\tau_M)+\beta(\tau_M)u^{-\tilde{\rho}}\right)^2\left(1+\frac{1}{ku}\right)du\right\}$$

$$\times\{1+o(1)\}$$

$$=(1-\gamma)^2\left(1+\frac{\bar{c}^2S_{cc}}{(S_{cc}+\tau_M)^2}\right)\{1+o(1)\},$$

Stellenbosch University https://scholar.sun.ac.za

and

$$
\begin{aligned}
&\frac{1}{k}\sum_{i=1}^{k}\sum_{j=i+1}^{k}\lambda_i(\tau_M)\,\lambda_j(\tau_M)\,Cov(\epsilon_i,\epsilon_j)\\
=&\frac{2(1-\gamma)}{1-2\gamma}\left\{\frac{1}{k}\sum_{i=1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{i}{k+1}\right)^{-\tilde{\rho}}\right)(i+1)\left(i^{-\gamma}-(i+1)^{-\gamma}\right)\right.\\
&\left.\times\sum_{j=i+1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}\right)(j+1)\left(j^{\gamma-1}-(j+1)^{\gamma-1}\right)\right\}\\
=&\frac{2(1-\gamma)^2\,\gamma}{1-2\gamma}\left\{\frac{1}{k^2}\sum_{i=1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{i}{k+1}\right)^{-\tilde{\rho}}\right)\frac{i+1}{k}\left(\frac{i}{k}\right)^{-\gamma-1}\right.\\
&\left.\times\sum_{j=i+1}^{k}\left(\alpha(\tau_M)+\beta(\tau_M)\left(\frac{j}{k+1}\right)^{-\tilde{\rho}}\right)\frac{j+1}{k}\left(\frac{j}{k}\right)^{\gamma-2}\right\}\\
=&\frac{2(1-\gamma)^2\,\gamma}{1-2\gamma}\left\{\int_{\frac{1}{k}}^{1}\left(\alpha(\tau_M)+\beta(\tau_M)u^{-\tilde{\rho}}\right)u^{-\gamma}\right.\\
&\left.\times\int_{u}^{1}\left(\alpha(\tau_M)+\beta(\tau_M)v^{-\tilde{\rho}}\right)v^{\gamma-1}\,dv\,du\right\}\{1+o(1)\}\\
=&\frac{2(1-\gamma)^2\,\gamma}{1-2\gamma}\left\{\int_{\frac{1}{k}}^{1}\left(\alpha(\tau_M)u^{-\gamma}+\beta(\tau_M)u^{-\tilde{\rho}-\gamma}\right)\right.\\
&\left.\times\left(\frac{\alpha(\tau_M)}{\gamma}\left(1-u^{\gamma}\right)+\frac{\beta(\tau_M)}{\gamma-\tilde{\rho}}\left(1-u^{\gamma-\tilde{\rho}}\right)\right)du\right\}\{1+o(1)\}\\
=&\frac{2(1-\gamma)\,\gamma}{1-2\gamma}\left\{\alpha^2(\tau_M)+\frac{\alpha(\tau_M)\beta(\tau_M)}{1-\tilde{\rho}}\left(1+\frac{1-\gamma}{1-\gamma-\tilde{\rho}}\right)\right.\\
&\left.+\frac{\beta^2(\tau_M)}{1-2\tilde{\rho}}\,\frac{1-\gamma}{1-\gamma-\tilde{\rho}}\right\}\{1+o(1)\}\\
=&\frac{2(1-\gamma)\,\gamma}{1-2\gamma}\left\{1-\frac{\tilde{\rho}}{1-\gamma-\tilde{\rho}}\,\frac{\bar{c}^2}{S_{cc}+\tau_M}+\frac{(1-\gamma)S_{cc}}{1-\gamma-\tilde{\rho}}\,\frac{\bar{c}^2}{(S_{cc}+\tau_M)^2}\right\}\{1+o(1)\}\,.
\end{aligned}
$$

□

Stellenbosch University https://scholar.sun.ac.za

Addendum B

Proposition *Assume $k, n \to \infty$ with $k/n \to 0$ such that $\sqrt{k} b_{n,k} \to M$ finite. Then under* (3.4)

$$\sqrt{k}(\hat{b}^{\bullet}_{n,k} - b_{n,k}) = O_p(1).$$

Proof In case of $\hat{b}^{\mathrm{CH}}_{n,k}$ we find that

$$\begin{aligned}
\sqrt{k}(\hat{b}^{\mathrm{CH}}_{n,k} - b_{n,k}) &= \sqrt{k}\left(\hat{\gamma}^{\mathrm{CH}}_{k}\hat{\beta}(n/k)^{\hat{\rho}} - \gamma\beta(n/k)^{\rho}\right) \\
&= \sqrt{k}\left(\hat{\gamma}^{\mathrm{CH}} - \gamma\right)\hat{\beta}(n/k)^{\hat{\rho}} + \sqrt{k}\left(\hat{\beta} - \beta\right)\gamma(n/k)^{\hat{\rho}} \\
&\quad + \sqrt{k}\left((n/k)^{\hat{\rho}} - (n/k)^{\rho}\right)\gamma\beta.
\end{aligned}$$

The result then follows from $\sqrt{k}\left(\hat{\gamma}^{\mathrm{CH}} - \gamma\right) = O_p(1)$, $\sqrt{k}\left(\hat{\beta} - \beta\right) = O_p(1)$, and $\sqrt{k}\left((n/k)^{\hat{\rho}} - (n/k)^{\rho}\right) = \sqrt{k}(n/k)^{\tilde{\rho}}\log(n/k)(\hat{\rho} - \rho)$ where $\log(n)(\hat{\rho} - \rho) = o_p(1)$ and $\tilde{\rho}$ is a random variable situated between $\hat{\rho}$ and ρ.

In case of $\hat{b}^{\mathrm{ridge}}_{n,k}$ we write

$$\begin{aligned}
&\sqrt{k}(\hat{b}^{\mathrm{ridge}}_{n,k} - b_{n,k}) \\
=&\sqrt{k}\left(\frac{1}{k}\sum_{j=1}^{k}\left[\frac{(d_{j,k} - \bar{d}_k)}{\frac{1}{k}\sum_{j=1}^{k}(d_{j,k} - \bar{d}_k)^2 + \hat{\tau}_k} - \frac{(c_{j,k} - \bar{c}_k)}{\frac{1}{k}\sum_{j=1}^{k}(c_{j,k} - \bar{c}_k)^2 + \tau}\right]Z_j\right) \\
&+ \sqrt{k}\left(\frac{1}{k}\sum_{j=1}^{k}\frac{(c_{j,k} - \bar{c}_k)Z_j}{\frac{1}{k}\sum_{j=1}^{k}(c_{j,k} - \bar{c}_k)^2 + \tau} - b_{n,k}\right). \qquad (4.12)
\end{aligned}$$

Using (3.5), up to $o_p(\sqrt{k}b_{n,k})$ the second term in (4.12) is given by

$$\gamma\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\frac{(c_{j,k} - \bar{c}_k)E_j}{\frac{1}{k}\sum_{j=1}^{k}(c_{j,k} - \bar{c}_k)^2 + \tau} + \sqrt{k}b_{n,k}\left(\frac{1}{k}\sum_{j=1}^{k}\frac{(c_{j,k} - \bar{c}_k)c_{j,k}E_j}{\frac{1}{k}\sum_{j=1}^{k}(c_{j,k} - \bar{c}_k)^2 + \tau} - 1\right),$$

which is bounded in probability since $\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(c_{j,k} - \bar{c}_k)E_j$ is asymptotically normal, $\frac{1}{k}\sum_{j=1}^{k}(c_{j,k} - \bar{c}_k)c_{j,k}E_j$ is bounded in probability, and $\sqrt{k}b_{n,k}$ is bounded by assumption.

Stellenbosch University https://scholar.sun.ac.za

The first term in (4.12) equals

$$\frac{\frac{1}{\sqrt{k}}\sum_{j=1}^{k}[(d_{j,k}-\bar{d}_k)-(c_{j,k}-\bar{c}_k)]Z_j}{\frac{1}{k}\sum_{j=1}^{k}(d_{j,k}-\bar{d}_k)^2+\hat{\tau}_k}$$

$$+\frac{\left(\frac{1}{k}\sum_{j=1}^{k}[(c_{j,k}-\bar{c}_k)^2-(d_{j,k}-\bar{d}_k)^2]+\tau-\hat{\tau}_k\right)\left(\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(c_{j,k}-\bar{c}_k)Z_j\right)}{\left(\frac{1}{k}\sum_{j=1}^{k}(c_{j,k}-\bar{c}_k)^2+\tau\right)\left(\frac{1}{k}\sum_{j=1}^{k}(d_{j,k}-\bar{d}_k)^2+\hat{\tau}_k\right)}.$$

Applying the mean value theorem to the function

$$\rho\mapsto\frac{1}{\sqrt{k}}\sum_{j=1}^{k}\left[(\frac{j}{k+1})^{-\rho}-\frac{1}{k}\sum_{l=1}^{k}(\frac{l}{k+1})^{-\rho})\right]Z_j$$

entails the boundedness in probability of the first term. Finally, since $\frac{1}{\sqrt{k}}\sum_{j=1}^{k}(c_{j,k}-\bar{c}_k)Z_j$ is asymptotically normal and $\hat{\tau}_k=O_p(1)$ the second term is also bounded in probability. (The case of $\hat{b}_{n,k}^{\text{ls}}$ follows similarly with $\tau=0$).

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Bibliography

Beirlant, J., G. Dierckx, Y. Goegebeur, and G. Matthys (1999), "Tail index estimation and an exponential regression model." *Extremes*, 2, 177–200.

Beirlant, J., G. Dierckx, and A. Guillou (2005), "Estimation of the extreme-value index and generalized quantile plots." *Bernoulli*, 11(6), 949–970.

Beirlant, J., G. Dierckx, A. Guillou, and C. Starica (2002), "On exponential representations of log-spacings of extreme order statistics." *Extremes*, 5, 157–180.

Beirlant, J., F. Figueiredo, M. I. Gomes, and B. Vandewalle (2008), "Improved reduced bias tail index and quantile estimators." *Journal of Statistical Planning and Inference*, 138, 1851–1870.

Beirlant, J., M. Gaonyalelwe, and A. Verster (2019), "Using shrinkage estimators to reduce bias and mse in estimation of heavy tails." *RevStat*, 17(1), 91–108.

Beirlant, J., Y. Goegebeur, J. Segers, and J. Teugels (2004), *Statistics of Extremes: Theory and Applications*. John Wiley & Sons Ltd, Chichester.

Bleistein, N. (1966), "Uniform asymptotic expansions of integrals with stationary point near algebraic singularity." *Communications on Pure and Applied Analysis*, 19(4), 353–370.

Boyd, S. and L. Vandenberghe (2004), *Convex Optimization*. Cambridge University Press, Cambridge.

Buitendag, S., J. Beirlant, and T. de Wet (2019), "Ridge regression estimators for the extreme value index." *Extremes*, 22(2), 271–292.

Butler, R. (2007), *Saddlepoint Approximations with Applications*. Cambridge University Press, Cambridge.

Caeiro, F., I. Gomes, and D. Pestana (2005), "Direct reduction of bias of the classical hill estimator." *RevStat*, 3(2), 113–136.

Stellenbosch University https://scholar.sun.ac.za

Cai, J.J., L. de Haan, and C. Zhou (2013), "Bias correction in extreme value statistics with index around zero." *Extremes*, 16(2), 173–201.

Chernozhukov, V., A. Galichon, M. Hallin, and M. Henry (2017), "Monge-kantorovich depth, quantiles, ranks, and signs." *Annals of Statistics*, 45, 223–256.

Daniels, H. E. (1954), "Saddlepoint approximation in statistics." *Annals of Mathematical Statistics*, 25, 631–650.

de Haan, L. and A. Ferreira (2006), *Extreme Value Theory: An Introduction*. Springer, New York.

de Haan, L. and H. Rootzén (1993), "On the estimation of high quantiles." *Journal of Statistical Planning and Inference*, 35, 1–13.

Dekkers, A. L. M., J. H. J. Einmahl, and L. de Haan (1989), "A moment estimator for the index of an extreme-value distribution." *Annals of Statistics*, 17, 1833–1855.

del Barrio, E., J. Cuesta-Albertos, M. Hallin, and C. Matran (2018), "Smooth cyclically monotone interpolation and empirical center-outward distribution functions." *Working Papers ECARES 2018-15, ULB - Université Libre de Bruxelles*.

del Barrio, E. and J.M. Loubes (2019), "Central limit theorems for empirical transportation cost in general dimension." *Annals of Probability*, 47, 926–951.

Ekeland, I., A. Galichon, and M. Henry (2012), "Comonotonic measures of multivariate risks." *Mathematical Finance*, 22, 109–132.

Embrechts, P., Klüppelberg, C., and T Mikosch (1997), *Modelling Extremal Events*. Springer, Berlin.

Feuerverger, A. and P. Hall (1999), "Estimating a tail exponent by modelling departure from a pareto distribution." *Annals of Statistics*, 27, 760–781.

Fisher, R. A. and L. H. C. Tippett (1928), "Limiting forms of the frequency distribution of the largest and smallest member of a sample." *Mathematical Proceedings of the Cambridge Philosophical Society*, 24(2), 180–190.

Fraga Alves, M.I., L. de Haan, and T. Lin (2003a), "Estimation of the parameter controlling the speed of convergence in extreme value theory." *Mathematical Methods of Statistics*, 12(2), 155–176.

Fraga Alves, M.I., M. I. Gomes, and L. de Haan (2003b), "A new class of semi-parametric estimators of the second order parameter." *Portugaliae Mathematica*, 60(2), 193–213.

Stellenbosch University https://scholar.sun.ac.za

Fréchet, M. (1927), "Sur la loi de probabilité de l'écart maximum." *Annales de la Société Polonaise de Mathématique*, 6(3), 93–116.

Gnedenko, B. V. (1943), "Sur la distribution limite du terme maximum d'une série aléatoire." *Annals of Mathematics*, 44(3), 423–453.

Gomes, M.I. and A. Guillou (2015), "Extreme value theory and statistics of univariate extremes: A review." *International Statistical Review*, 83(2), 263–292.

Gomes, M.I. and M.J. Martins (2002), "Asymptotically unbiased estimators of the tail index based on external estimation of the second order parameter." *Extremes*, 5, 387–414.

Gomes, M.I., M.J. Martins, and M. Neves (2007), "Improving second order reduced bias extreme value index estimation." *RevStat*, 5, 177–207.

Gomes, M.I. and D. Pestana (2007), "A sturdy reduced bias extreme quantile (VaR) estimator." *Journal of the American Statistical Association*, 102(477), 280–292.

Guillou, A. and P. Hall (2001), "A diagnostic for selecting the threshold in extreme value analysis." *Journal of the Royal Statistical Society B*, 63, 293–305.

Gushchin, A.A. and D.A. Borzykh (2018), "Integrated quantile functions: properties and application." *Modern Stochastics: Theory and Applications*, 4(4), 285–314.

Haeusler, E and S. Segers (2007), "Assessing confidence intervals for the tail index by edgeworth expansions for the hill estimator." *Bernoulli*, 13(1), 175–194.

Hall, P. (1982), "On some simple estimates of an exponent of regular variation." *Journal of the Royal Statistical Society B*, 44, 37–42.

Hallin, M. (2017), "Distribution and quantile functions, ranks and signs in $\mathbb{R}^d$: a measure transportation approach." *ECARES Working paper*.

Hastie, T., R. Tibshirani, and J. H. Friedman (2008), *The Elements of Statistical Learning*. Springer-Verlag, New York.

Hill, B. M. (1975), "A simple general approach to inference about the tail of a distribution." *Annals of Statistics*, 3(5), 1163–1174.

Hoerl, A.E. and R.W. Kennard (1970), "Ridge regression: biased estimation for nonorthogonal problems." *Technometrics*, 12, 55–67.

Stellenbosch University https://scholar.sun.ac.za

Hosking, J. R. M. and J. R. Wallis (1987), "Parameter and quantile estimation for the generalized pareto distribution." *Technometrics*, 29, 339–349.

Hosking, J. R. M., J. R. Wallis, and E. F. Wood (1985), "Estimation of the generalized extreme-value distribution by the method of probability-weighted moments." *Technometrics*, 27, 251–261.

Jensen, J. L. (1995), *Saddlepoint Approximations*. Clarendon Press, Oxford.

Kim, M. and S. Lee (2017), "Estimation of the tail exponent of multivariate regular variation." *Annals of the Institute of Statistical Mathematics*, 69, 945–968.

Kusuoka, S. (2001), "On law invariant coherent risk measures." *Advances in Mathematical Economics*, 3, 83–95.

Lugannani, R. and S Rice (1980), "Saddle point approximation for the distribution of the sum of independent random variables." *Advances in Applied Probability*, 12(2), 475–490.

Matthys, G. and J. Beirlant (2000), "Adaptive threshold selection in tail index estimation." In *Extremes and Integrated Risk Management* (P. Embrechets, ed.), 37–49, Risk Books, London.

Matthys, G., E. Delafosse, A. Guillou, and J. Beirlant (2004), "Estimating catastrophic quantile levels for heavy-tailed distributions." *Insurance: Mathematics and Economics*, 7, 271–300.

Pickands, J. (1975), "Statistical inference using extreme order statistics." *Annals of Statistics*, 3, 119–131.

Prescott, P. and A. T. Walden (1980), "Maximum likelihood estimation of the parameters of the generalized extreme-value distribution." *Biometrika*, 67, 723–724.

Prescott, P. and A. T. Walden (1983), "Maximum likelihood estimation of the parameters of the three-parameter generalized extreme value distribution from censored samples." *Journal of Statistical Computation and Simulation*, 16, 241–250.

Rockaffelar, R.T. (1996), "Characterization of the subdifferential of convex functions." *Pacific Journal of Mathematics*, 17, 497–510.

Rüschendorf, L. (2006), "Law invariant convex risk measures for portfolio vectors." *Statistics & Decisions*, 24, 97–108.

Smith, R.L. (1985), "Maximum likelihood estimation in a class of nonregular cases." *Biometrika*, 72, 67–90.

Stellenbosch University https://scholar.sun.ac.za

Smith, R.L. (1987), "Estimating tails of probability distributions." *Annals of Statistics*, 15, 1174–1207.

Villani, C. (2009), *Optimal Transport: Old and New*. Springer-Verlag, Heidelberg.

von Mises, R. (1936), "La distribution de la plus grande de n valeurs." *Revue Mathematique Union Interbalcanique*, 1, 141–160.

Weissman, I. (1978), "Estimation of parameters and large quantiles based on the k largest observations." *Journal of the American Statistical Association*, 73, 812–815.

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Publications

Buitendag, S., Beirlant, J. and de Wet, T. (2019). Ridge regression estimators for the extreme value index. *Extremes*, 22(2), 271-292.

Buitendag, S., Beirlant, J. and de Wet, T. (2019). Confidence intervals for extreme Pareto-type quantiles. To appear in *Scandinavian Journal of Statistics*, available on https://doi.org/10.1111/sjos.12396.

Beirlant, J., Buitendag, S., del Barrio, T. and Hallin, M. (2019). Center-outward quantiles and the measurement of multivariate risk. Submitted to *Insurance: Mathematics and Economics*, available on https://arxiv.org/abs/1912.04924.

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

Stellenbosch University https://scholar.sun.ac.za

FACULTY OF SCIENCE
DEPARTMENT OF MATHEMATICS
Celestijnenlaan 200B, 3001 Leuven
wis.kuleuven.be

FACULTY OF ECONOMIC AND MANAGEMENT SCIENCES
DEPARTMENT OF STATISTICS AND ACTUARIAL SCIENCE
47 Victoria Street, Stellenbosch, 7600
www.sun.ac.za/english/faculty/economy/statistics